CMIP6 Models Underestimate Arctic Sea Ice Loss during the Early Twentieth-Century Warming, despite Simulating Large Low-Frequency Sea Ice Variability

ELENA BIANCO,[a,b,c] EDWARD BLANCHARD-WRIGGLESWORTH,[c] STEFANO MATERIA,[a,d] PAOLO RUGGIERI,[e] DOROTEACIRO IOVINO,[a] AND SIMONA MASINA[a]

[a] *Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy*
[b] *Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy*
[c] *Department of Atmospheric Sciences, University of Washington, Seattle, Washington*
[d] *Barcelona Supercomputing Center, Barcelona, Spain*
[e] *Department of Physics and Astronomy, University of Bologna, Bologna, Italy*

(Manuscript received 24 October 2023, in final form 11 August 2024, accepted 3 September 2024)

ABSTRACT: The variability of Arctic sea ice extent (SIE) on interannual and multidecadal time scales is examined in 29 models with historical forcing participating in phase 6 of the Coupled Model Intercomparison Project (CMIP6) and in twentieth-century sea ice reconstructions. Results show that during the historical period with low external forcing (1850–1919), CMIP6 models display relatively good agreement in their representation of interannual sea ice variability (IVSIE) but exhibit pronounced intermodel spread in multidecadal sea ice variability (MVSIE), which is overestimated with respect to sea ice reconstructions and is dominated by model uncertainty in sea ice simulation in the subpolar North Atlantic. We find that this is associated with differences in models' sensitivity to Northern Hemispheric sea surface temperatures (SSTs). Additionally, we show that while CMIP6 models are generally capable of simulating multidecadal changes in Arctic sea ice from the mid-twentieth century to present day, they tend to underestimate the observed sea ice decline during the early twentieth-century warming (ETCW; 1915–45). These results suggest the need for an improved characterization of the sea ice response to multidecadal climate variability in order to address the sources of model bias and reduce the uncertainty in future projections arising from intermodel spread.

SIGNIFICANCE STATEMENT: The credibility of Arctic sea ice predictions depends on whether climate models are capable of reproducing changes in the past climate, including patterns of sea ice variability which can mask or amplify the response to global warming. This study aims to better understand how latest-generation global climate models simulate interannual and multidecadal variability of Arctic sea ice relative to available observations. We find that models differ in their representation of multidecadal sea ice variability, which is overall larger than in observations. Additionally, models underestimate the sea ice decline during the period of observed warming between 1915 and 1945. Our results suggest that, to achieve better predictions of Arctic sea ice, the realism of low-frequency sea ice variability in models should be improved.

KEYWORDS: Arctic; Sea ice; Climate models; Interannual variability; Multidecadal variability

1. Introduction

The rapid decline of Arctic sea ice since the beginning of the satellite era (e.g., Cavalieri et al. 2003; Comiso et al. 2008) has created a growing need for accurate predictions of future Arctic climate changes, including estimates of the timing of ice-free conditions in summer (e.g., Jahn et al. 2024), and for enhanced understanding on the variability of sea ice. Global climate models (GCMs) are the most comprehensive tool for analyzing trends and variability of Arctic sea ice, both historically and in future projections. The Coupled Model Intercomparison Project (CMIP) provides a valuable framework for evaluating and improving the representation of sea ice in latest-generation GCMs. Over the years, significant advancements have been made in model formulation and realism, with consequent improvements in terms of sea ice extent and edge (Davy and Outten 2020) and better representation of spatial patterns (Long et al. 2021). Recent assessments showed that over the central Arctic, CMIP6 models outperform the ERA5 reanalysis in capturing the observed warming trend (Tian et al. 2024).

Despite these promising results, CMIP6 models still exhibit large intermodel spread, and discrepancies with observations persist (Shen et al. 2021; Notz and SIMIP Community 2020; Watts et al. 2021). Several factors can contribute to the mismatch between simulated and observed Arctic changes, including a biased representation of sea ice and other climate system components (e.g., Topál et al. 2020; Khosravi et al. 2022; Massonnet et al. 2018). A number of studies have pointed to unrealistic sea ice sensitivity to warming as an

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-23-0647.s1.

Corresponding author: Elena Bianco, elena.bianco@cmcc.it

DOI: 10.1175/JCLI-D-23-0647.1

important reason for why model simulations often differ from observations (Winton 2011; Mahlstein and Knutti 2012; Rosenblum and Eisenman 2017). Specifically, Notz and SIMIP Community (2020) showed that the majority of CMIP6 models simulate insufficient sea ice sensitivity to global mean surface temperatures. At short weather time scales (<20 days), models tend to simulate too little sea ice variability compared to observations (Blanchard-Wrigglesworth et al. 2021).

Besides model formulation, substantial uncertainty arises from internal climate variability associated with atmospheric (e.g., Olonscheck et al. 2019) and oceanic (e.g., Li et al. 2017) drivers. The observational record represents only one plausible trajectory of sea ice response to changes in anthropogenic forcing and internal climate variability. Considerable effort has gone toward assessing the relative contributions of these two factors to recent Arctic changes (e.g., Ye and Messori 2021; Shen et al. 2021), with many studies concluding that a substantial portion (20%–50%) of the observed sea ice trend is attributable to natural variability (Kay et al. 2011; Ding et al. 2017, 2019; England et al. 2019; Dörr et al. 2023).

The physical processes that influence sea ice variability can vary considerably across time scales. For instance, sea ice variability on synoptic to subseasonal time scales in GCMs has been linked to local wind and ocean wave characteristics (Blanchard-Wrigglesworth et al. 2021), while interannual and interdecadal changes have been associated with tropical–arctic teleconnections (e.g., Baxter et al. 2019; Clancy et al. 2021), though the robustness of the teleconnections is unclear (Bonan and Blanchard-Wrigglesworth 2020). On decadal and multidecadal time scales, sea ice variability occurs in response to thermodynamic processes associated with atmospheric and oceanic forcing (Bitz et al. 1996), particularly ocean heat transport from the Atlantic Ocean (Zhang 2015; Li et al. 2017; Yeager et al. 2015) and AMOC variability (Liu and Fedorov 2022). However, our understanding of multidecadal sea ice variability remains limited, largely because the insufficient temporal coverage of sea ice observations has so far limited characterization of these time scales. Moreover, the satellite record provides a single realization of the climate system that is also undergoing rapid changes, which complicates the disentanglement of forced and natural contributions (England et al. 2019; Ding et al. 2019).

Observations of Arctic temperatures throughout the twentieth century have revealed patterns of multidecadal variability characterized by an early period of warming over 1915–45, cooling over 1945–75, and the modern, ongoing period of warming since 1975 (Polyakov et al. 2003; Bengtsson et al. 2004). The first period of warming is known as the early twentieth-century warming (ETCW) and was characterized by abrupt temperature increases in the Arctic with comparable warming trends to the modern satellite-era warming, albeit of shorter duration (Bengtsson et al. 2004; Johannessen et al. 2004; Hegerl et al. 2018). Evidence of significant sea ice decline during the ETCW has been documented in twentieth-century sea ice extent reconstructions (Brennan et al. 2020), with an estimated loss of Arctic sea ice area of 1.5 million km^2 over 1915–45. The driving mechanisms of the ETCW are still a subject of debate. Several previous studies concluded that the warming originated from an internally generated superposition of Northern Hemispheric climate anomalies (e.g., Overland and Wang 2005; Wood and Overland 2010; Johannessen et al. 2004; Beitsch et al. 2014), enhanced by positive feedbacks in the Arctic (Bokuchava and Semenov 2021). Other studies (Fyfe et al. 2013; Hegerl et al. 2018; Aizawa et al. 2021) have suggested a contribution of external forcing agents, including positive radiative forcing from black carbon and decline in late 19th century/early twentieth century volcanism, though relative roles are uncertain. On the other hand, the relative cooling and associated sea ice recovery that followed the ETCW in the mid-twentieth century have been ascribed to increased emissions of anthropogenic aerosols (England et al. 2021; Aizawa et al. 2022). The magnitude and potentially recurrent nature of the ETCW highlight the importance of characterizing the sensitivity and low-frequency variability of Arctic sea ice in GCMs, in order to gain confidence in their representation of the present and future Arctic climate. Moreover, earlier generations of climate models (CMIP3 and CMIP5) have consistently failed to capture the magnitude and duration of the ETCW (Fyfe et al. 2013; Wang et al. 2007), which motivates us to investigate whether CMIP6 models show any improvement in replicating this event.

In this study, we examine changes and variability of Arctic sea ice in 29 CMIP6 historical simulations (1850–2014) using available observations for comparison. By analyzing a large subset of CMIP6 models with all their available ensemble members, we illustrate the range of simulated sea ice variability between models, taking into account differences in model sensitivity as well as the uncertainty originating from internal climate variability.

Specifically, we address the following questions: 1) How do CMIP6 models compare in simulating interannual and multidecadal sea ice variability over the historical period? 2) How is sea ice variability on these time scales linked to global ocean and air temperatures? And finally, 3) do CMIP6 models capture multidecadal sea ice changes and in particular the decline associated with the ETCW?

2. Data and methods

We analyze fully coupled, free-running simulations with historical forcing (1850–2014) from 29 CMIP6 models originating from 19 institutions, using all their available ensemble members. All but six CMIP6 models (CMCC-CM2-HR4, CMCC-ESM2, CNRM-CM6-1-HR, GFDL-CM4, SAM0-UNICON, and TaiESM1) include multiple ensemble members ranging up to 65. A brief description of the CMIP6 subset employed in this study is provided in Table 1.

In addition to CMIP6 historical simulations, we consider annual Arctic sea ice extent (SIE) reconstructions since 1850 from Brennan et al. (2020, hereafter B20SIE). The product relies on an ensemble Kalman filter data assimilation approach, combining model data from fully forced Last Millennium simulations covering the period 850–1849 with instrumental temperature observations. In particular, we consider two B20SIE ensembles generated with the MPI-ESM-P model assimilating

TABLE 1. List of the 29 CMIP6 models analyzed in this study with their ensemble size, native nominal ocean grid resolution, and sea ice model component.

Model name	Modeling center	Country	Ensemble members[a]	Native nominal ocean resolution (km)	Sea ice model
ACCESS-CM2	CSIRO-ARCCSS	Australia	3	100	CICE5.1.2
ACCESS-ESM1-5	CSIRO	Australia	5	100	CICE4.1
BCC-ESM1	BCC	China	3	50	SIS2
CAMS-CSM1-0	CAMS	China	3	100	SIS1.0
CESM2	NCAR	United States	11	100	CICE5.1
CESM2-WACCM	NCAR	United States	3	100	CICE5.1
CMCC-CM2-HR4	CMCC	Italy	1	25	CICE4.0
CMCC-ESM2	CMCC	Italy	1	100	CICE4.0
CNRM-CM6-1	CNRM-CERFACS	France	21	100	Gelato6.1
CNRM-CM6-1-HR	CNRM-CERFACS	France	1	25	Gelato6.1
CNRM-ESM2-1	CNRM-CERFACS	France	6	100	Gelato6.1
CanESM5	CCCma	Canada	65	100	LIM2
CanESM5-CanOE	CCCma	Canada	3	100	LIM2
E3SM-1-0	E3SM-Project	United States	5	50	MPAS-Seaice
EC-Earth3	EC-Earth-Consortium	Europe	23	100	LIM3
GFDL-CM4	NOAA-GFDL	United States	1	25	GFDL-SIM4p25
GFDL-ESM4	NOAA-GFDL	United States	3	50	GFDL-SIM4p2
HadGEM3-GC31-LL	MOHC NERC	United Kingdom	4	100	CICE-HadGEM3-GSI8
IPSL-CM6A-LR	IPSL	France	32	100	NEMO-LIM3
MIROC-ES2L	MIROC	Japan	31	100	COCO4.9
MIROC6	MIROC	Japan	50	100	COCO4.9
MPI-ESM1-2-LR	MPI-M AWI	Germany	10	250	Unnamed[b]
MRI-ESM2-0	MRI	Japan	6	100	MRI-COM4.4
NASA-GISS-E2-1-H	NASA-GISS	United States	25	100	GISS SI
NCC-NorCPM1	NCC	Norway	30	100	CICE4
NCC-NorESM2-MM	NCC	Norway	2	100	CICE
SAM0-UNICON	SNU	Korea	1	100	CICE4.0
TaiESM1	AS-RCEC	Taiwan	1	100	CICE4
UKESM1-0-LL	MOHC NERC	United Kingdom	16	100	CICE-HadGEM3-GSI8

[a] Refers to the number of available ensemble members for the variable "siconc." This may vary for other variables.
[b] Thermodynamic (Semtner zero-layer) dynamic (Hibler 79) sea ice model.

HadCRUT4 temperatures (B20SIE MPI-HadCRUT4) and the CCSM4 model combined with Berkeley Earth observations (B20SIE CCSM4-BE). Each B20SIE ensemble consists of five independent iterations that each uses a 200-member prior ensemble, for a total of 1000 members. For additional comparison, we use sea ice observations from the National Snow and Ice Data Center (NSIDC) Sea Ice Index, version 3.0 (Fetterer et al. 2017; https://nsidc.org/data/g02135/versions/3) for the satellite period (1979–2020).

To assess sea ice variability during a time of low external forcing, we focus on 1850–1919, a period with minimal anthropogenic influence and little global warming, when no statistically significant trend in ensemble mean SIE can be detected.

We compute Arctic SIE by integrating the area of sea ice above 15% concentration (SIC) between 40° and 90°N on each model's native grid. Time series of annual mean SIE anomalies are partitioned into multidecadal (MVSIE) and interannual (IVSIE) time series. MVSIE is obtained by applying a Butterworth low-pass filter with a 10-yr cutoff frequency. IVSIE is then derived by subtracting the filtered time series from annual SIE anomalies to retain interannual and subinterannual frequencies of variability. By using a 10-yr cutoff frequency, the influence of climate variability modes that operate at interannual time scales, such as El Niño–Southern Oscillation (ENSO; 2–7 year periodicity), is separated from low-frequency modes, namely, the Atlantic multidecadal oscillation (AMO) and Pacific decadal oscillation (PDO), which affect Arctic sea ice on multidecadal time scales (e.g., Zhang 2015). We apply the same filtering method to annual anomalies of sea surface temperature (SST) and near-surface air temperature (SAT) for the same period in order to quantify models' sensitivity to global temperature patterns and identify large-scale sources of sea ice variability.

The last 100 years of historical simulations (1915–2014) are characterized by periods of warming and relative cooling under greater external anthropogenic forcing and are therefore analyzed separately. In this study, our main aim is to assess whether CMIP6 models simulate a realistic ETCW sea ice decline with respect to the B20SIE reconstructions. We additionally examine SIE and SAT trends in the 29 CMIP6 models along with the B20SIE reconstructions and satellite observations for the modern warming period (1979–2014) and the intermediate period of observed Arctic cooling following the ETCW (1946–78). To accomplish this, we obtain time

series of global and Arctic (70°–90°N) near-surface surface temperature from gridded observations of Met Office Hadley Centre/Climatic Research Unit global surface temperature dataset analysis, version 5.0.1.0 (HadCRUT5) (Morice et al. 2021; https://crudata.uea.ac.uk/cru/data/temperature/). HadCRUT5 observations are expressed as anomalies relative to the 1961–90 period and rely on a statistical infilling method to improve the representation of sparsely observed regions. For the ETCW, we also utilize the HadCRUT5 noninfilled dataset, which adopts the gridding method of HadCRUT4 (Morice et al. 2012), to quantify temperature changes where measurements are available.

3. Results

a. *Interannual and multidecadal sea ice variability during 1850–1919*

We first examine the variability of MVSIE and IVSIE during the first 70 years of historical simulations (1850–1919). This period is characterized by a nonsignificant trend in simulated sea ice and minimal anthropogenic forcing, and thus internal climate variability is expected to be the dominant factor influencing sea ice fluctuations.

Figure 1 illustrates the variability of MVSIE and IVSIE over 1850–1919 (expressed as standard deviations of SIE anomalies, σ_{MVSIE} and σ_{IVSIE}) in 29 CMIP6 models and its relation to the mean sea ice state. For comparison, we also show σ_{MVSIE} and σ_{IVSIE} in B20SIE reconstructions covering the same period. The σ_{MVSIE} in CMIP6 models is generally overestimated with respect to the B20SIE reconstructions (with a multimodel mean σ_{MVSIE} of 0.34 million km^2 compared to 0.13–0.2 million km^2 in B20SIE) and exhibits substantially larger uncertainty than σ_{IVSIE}, both in terms of intermodel discrepancies and ensemble spread within individual GCMs. The spread in MVSIE is almost fivefold between the most and least variable models (0.77 million km^2 in EC-Earth3 versus 0.16 million km^2 in MIROC6, as further illustrated by the time series in Fig. S1 in the online supplemental material). Generally, the simulated variability of MVSIE is about twice as large as that of IVSIE (Fig. S2). Overall, CESM2-WACCM and GFDL-ESM4 exhibit the closest agreement with the B20SIE reconstructions for both MVSIE and IVSIE variabilities, although this could in part be influenced by their small ensemble size. From Fig. 1a, it is apparent that the ensemble mean σ_{MVSIE} in EC-Earth3 is overestimated compared to the other models. Conversely, MIROC6 and MIROC-ES2L simulate the smallest SIE variability among the CMIP6 subset on both time scales. We find that these models also exhibit biased low mean SIE during 1850–1919. Across the CMIP6 multimodel ensemble, models with greater mean state sea ice cover tend to simulate enhanced variability; however, the relationships are weak ($r = 0.36$ for IVSIE and $r = 0.24$ for MVSIE, which is not significant at the 95% threshold; Figs. 1c,d). This indicates that while the models' ability to simulate interannual sea ice variability is in part influenced by their representation of the mean sea ice state, this factor alone is not sufficient to explain the pronounced model spread and bias with respect to B20SIE that is observed at multidecadal time scales.

Next, we analyze the spatial distribution of multidecadal and interannual SIE variability in CMIP6 models to diagnose the regions that drive sea ice variability and where models agree or disagree. Figure 2 shows the CMIP6 multimodel mean (MMM) and intermodel spread σ of the correlation between SIE and SIC on multidecadal and interannual time scales (where MVSIC and IVSIC are defined by applying the same filtering method described in section 2 to SIC) in models with at least 10 ensemble members. MVSIE variability in the MMM is driven by MVSIC in the Barents and the subarctic seas, especially in the North Atlantic, likely due to the influence of low-frequency variations in North Atlantic Ocean temperatures (further discussed in section 3b).

On the other hand, IVSIE variability is driven by IVSIC variability in the Barents Sea and along the Siberian Arctic but less so by IVSIC variability in the subarctic North Atlantic (note the difference to the east and south of Greenland between Figs. 2a,b). Consistent with the pattern of intermodel spread of pan-Arctic SIE variability (Fig. 1), individual models display large uncertainty in the representation of MVSIE (Figs. 2c and 3, which shows individual models' r[MVSIE, MVSIC] but relatively good agreement in simulating the regional drivers of IVSIE (Figs. 2d and 4, which shows individual models' r[IVSIE, IVSIC]. On multidecadal time scales, models show the largest disagreement in the contribution of MVSIC variability to MVSIE in the Atlantic Arctic sector, particularly in the Labrador and Greenland seas, whereas practically all models agree in the large contribution of MVSIC in the Barents Sea to MVSIE (note the low σ values in the Barents Sea in Fig. 2c).

b. *Simulated sea ice sensitivity to global temperatures*

Building on the understanding that Arctic sea ice on interannual and longer time scales is strongly coupled to global temperatures (e.g., Mahlstein and Knutti 2012; Winton 2011; Gregory et al. 2002), we explore the links between SIE and global SST to assess whether the intermodel spread in SIE variability is attributable to differences in models' sensitivity to large-scale climate variability and to investigate the differences in the patterns of Arctic SIE–global climate coupling across multidecadal and interannual time scales. Figure 5 shows the MMM regression of MVSIE and IVSIE on standardized global SST during 1850–1919. Regression maps for individual models can be found in the supplemental material. Temperature anomalies are standardized prior to computing the regressions to facilitate intermodel comparison.

MVSIE in the MMM is tightly coupled to multidecadal sea surface temperature (MVSST) in the Northern Hemisphere, as reflected by the significant regression coefficients extending from the Arctic Ocean to the equatorial Atlantic and Pacific Oceans (Fig. 5a). Conversely, the link between SIE and SST on interannual time scales is primarily limited to the Arctic Ocean north of 60°N (Fig. 5b), illustrating how in general MVSIE is coupled to global climate, whereas IVSIE appears to be isolated to Arctic climate variability. The patterns of SIE–SAT coupling on both time scales (Fig. S5) are largely

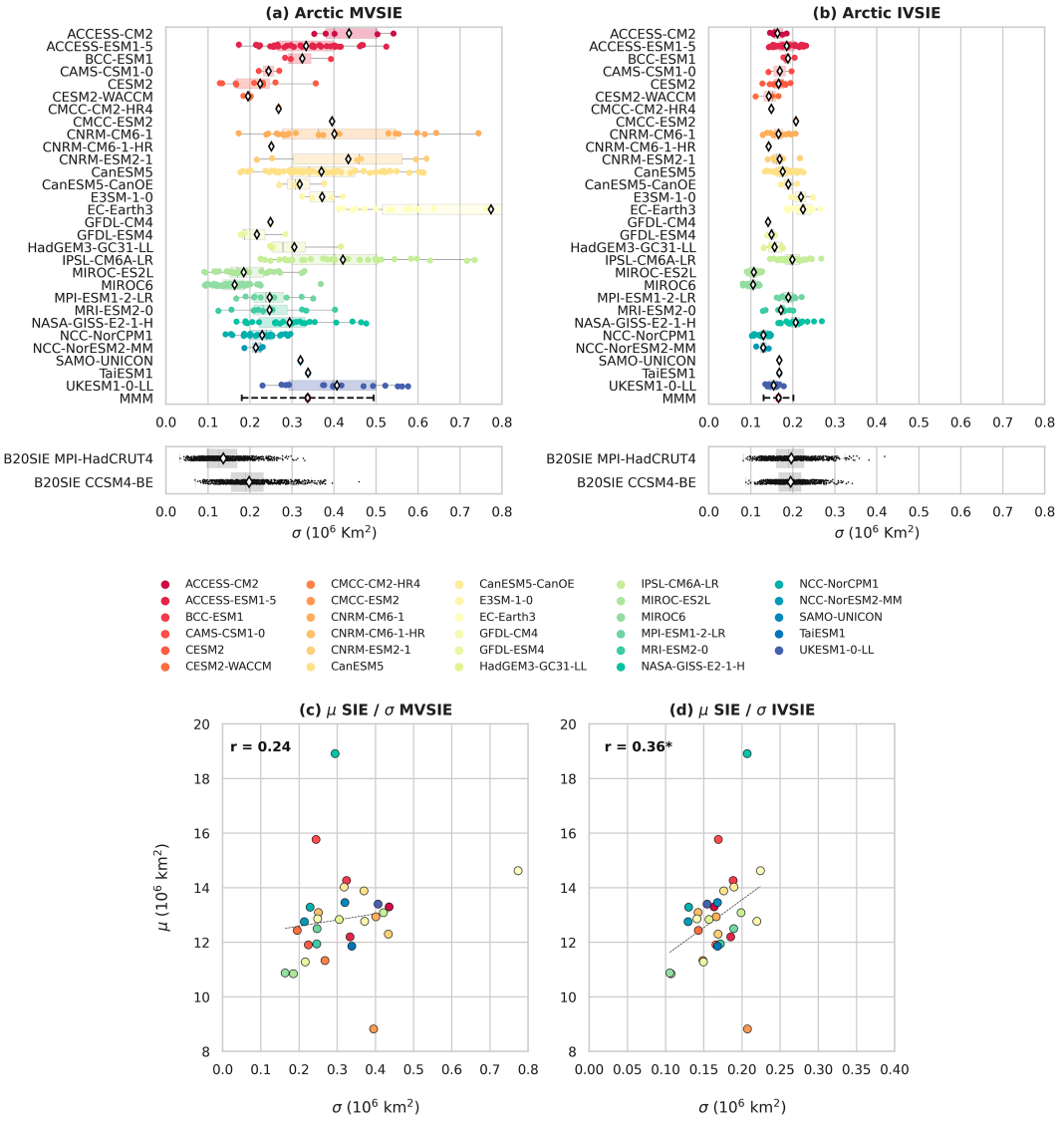

FIG. 1. Standard deviations of (a) multidecadal and (b) interannual Arctic SIE anomalies (1850–1919) in 29 CMIP6 models and B20SIE reconstructions with two model priors [lower panels in (a) and (b)]. Single dots indicate individual ensemble members, shaded boxes indicate the interquartile range, and white diamonds mark the ensemble mean for each model/product. The CMIP6 MMM in (a) and (b) denotes the average across the ensemble mean SIE standard deviations of models with at least 10 ensemble members. Dashed lines in the MMM mark one standard deviation from the mean. The ensemble of dots in each B20SIE reconstruction corresponds to the 1000-member model prior ensemble and represents the uncertainty due to sampling error. (c),(d) The relation between (c) multidecadal and (d) interannual ensemble mean SIE standard deviations and mean sea ice state (1850–1919) in the 29 CMIP6 models.

consistent with those of SIE and SST, though with lower regression coefficients in the Arctic due to the larger internal variability of atmospheric temperatures relative to SST, which is constrained close to the freezing point in regions of sea ice cover. We note that the coupling between Atlantic SST and Arctic SIE that emerges from the CMIP6 MMM (Fig. 5a) is also consistent with observations (Fig. S6, showing the regression of annual B20SIE anomalies on global SST observations from HadSST4).

Individual models display considerable differences in their representation of SIE sensitivity to global SST (Figs. S3 and S4). MIROC6 and MIROC-ES2L stand out because of remarkably low regression coefficients outside the Arctic on both time scales, suggesting insufficient sea ice sensitivity to changes in Northern Hemispheric SST. This in part explains the smaller MVSIE and IVSIE variability that is observed in these models (Fig. 1), particularly in the Atlantic Arctic (Fig. 3). Conversely, EC-Earth3 and E3SM-1-0 display unrealistic high

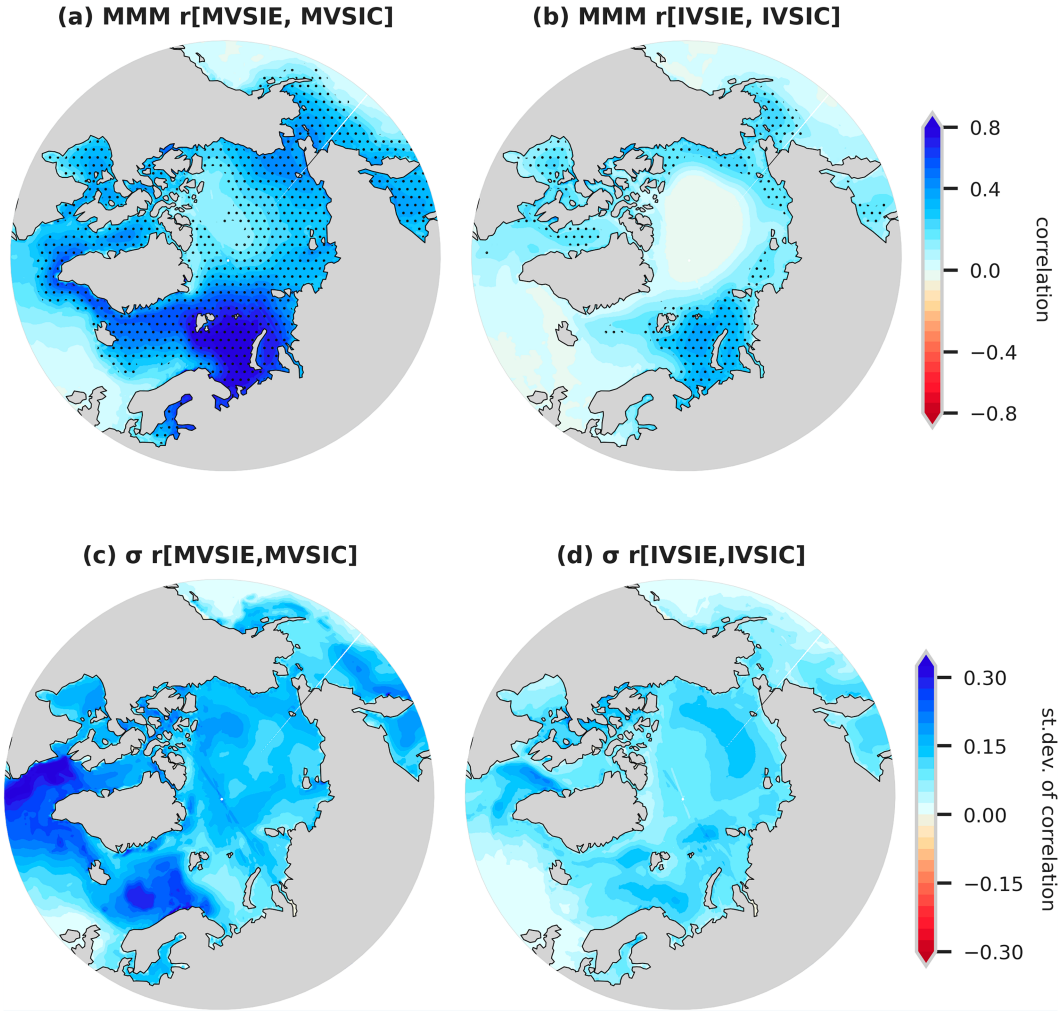

FIG. 2. (a),(b) Multimodel mean and (c),(d) standard deviation of the correlation between Arctic SIE and grid-point SIC during 1850–1919 for (a),(c) multidecadal and (b),(d) interannual timescales. Only models with at least 10 members are included in the multimodel ensemble. The correlations are Fisher-*Z*-transformed prior to computing the multimodel mean. Stippling in (a) and (b) denotes 95% significance according to a Student's *t* test.

and widespread sea ice sensitivity to multidecadal SST, though this feature does not emerge on interannual time scales. For EC-Earth3, the biased high sensitivity of MVSIE to global SST is consistent with the overestimated MVSIE variability (Fig. 1a).

Furthermore, we find that models that simulate lower MVSIE variability in the Greenland and Labrador Seas relative to the CMIP6 MMM (including CESM2, CESM2-WACCM, MIROC6, MIROC-ES2L, and MPI-ESM1-2-LR; Fig. 3) also display weak regressions with MVSST in the Northern Hemisphere (Fig. S3). We examine whether these differences originate from sensitivity biases or rather a misrepresentation of Northern Hemispheric SST variability. Figure 6 illustrates the links between the multidecadal variability of Arctic MVSIE (σ_{MVSIE}) and Northern Hemispheric SST ($\sigma_{\mathrm{MVSST_{NH}}}$) during 1850–1919. Models with especially low (high) MVSIE sensitivity to $\mathrm{MVSST_{NH}}$ (Fig. 6a, Fig. S3) tend to have smaller (larger) σ_{MVSIE} (Fig. 6a) and $\sigma_{\mathrm{MVSST_{NH}}}$ (Fig. 6b), suggesting that the strength of MVSIE–MVSST coupling is largely explained by models' representation of Northern Hemispheric SST variability (Fig. 6c). For instance, EC-Earth3's overestimated σ_{MVSIE} is due to the combined contributions of high sensitivity (Figs. 6a,b) and a generally overestimated Northern Hemispheric climate variability (Fig. 6c, Fig. S7a). It is worth noting that the correlation between MVSIE and $\mathrm{MVSST_{NH}}$ in CMIP6 models peaks when $\mathrm{MVSST_{NH}}$ leads MVSIE by 1 year and is significant up to a lag of 3 years (Fig. S8a), suggesting that the main direction of causality is temperature variability forcing sea ice variability.

c. *Multidecadal sea ice changes and the ETCW*

In the following section, we shift our focus to the final century of CMIP6 historical simulations, covering the years 1915–2014. The observational record of Arctic temperatures

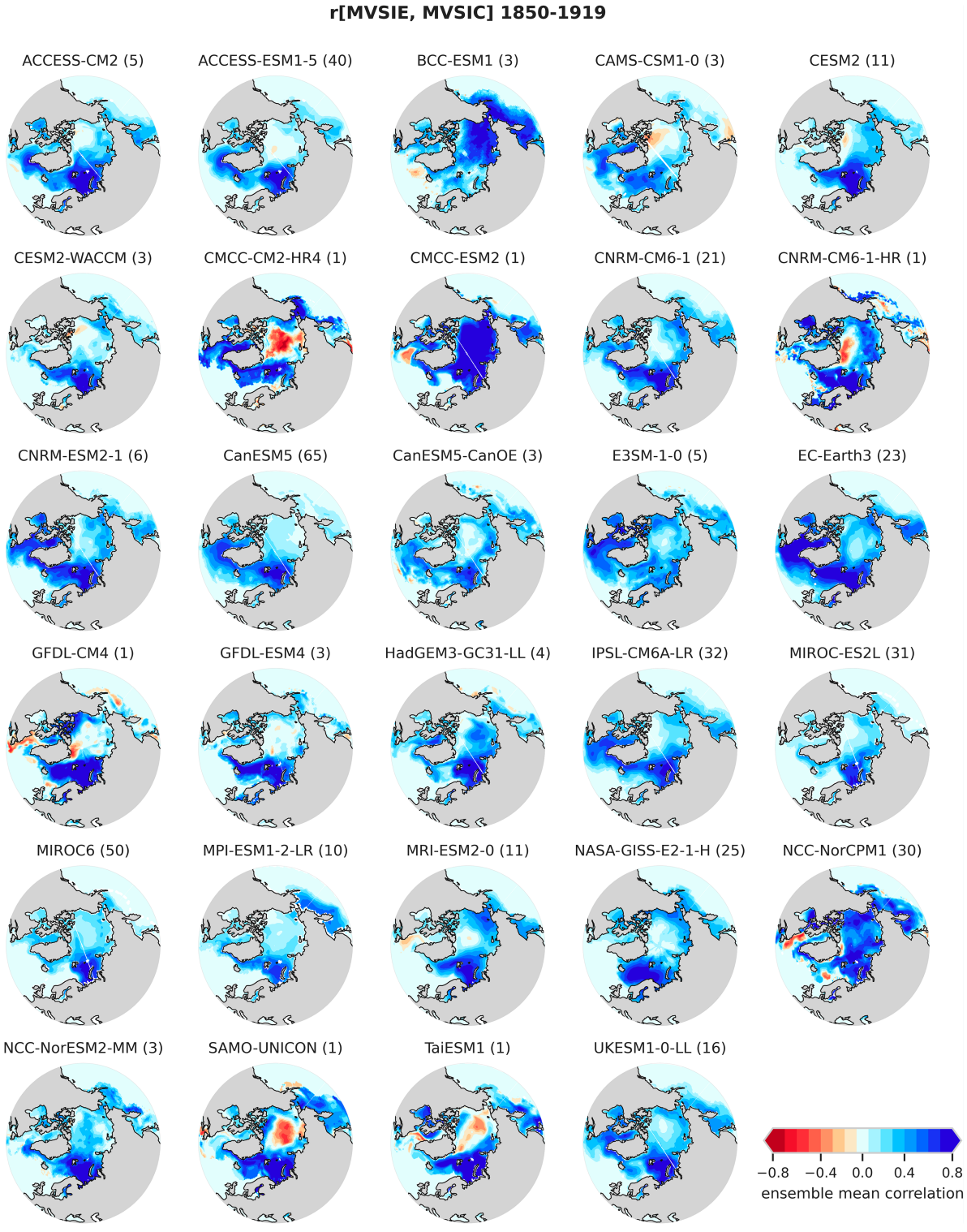

FIG. 3. Ensemble mean correlation between multidecadal Arctic SIE and gridpoint SIC in 29 CMIP6 models. Values in parentheses refer to the number of ensemble members. For any given model, the ensemble mean correlation is computed as the average correlation across the model's ensemble members after Fisher's Z transformation.

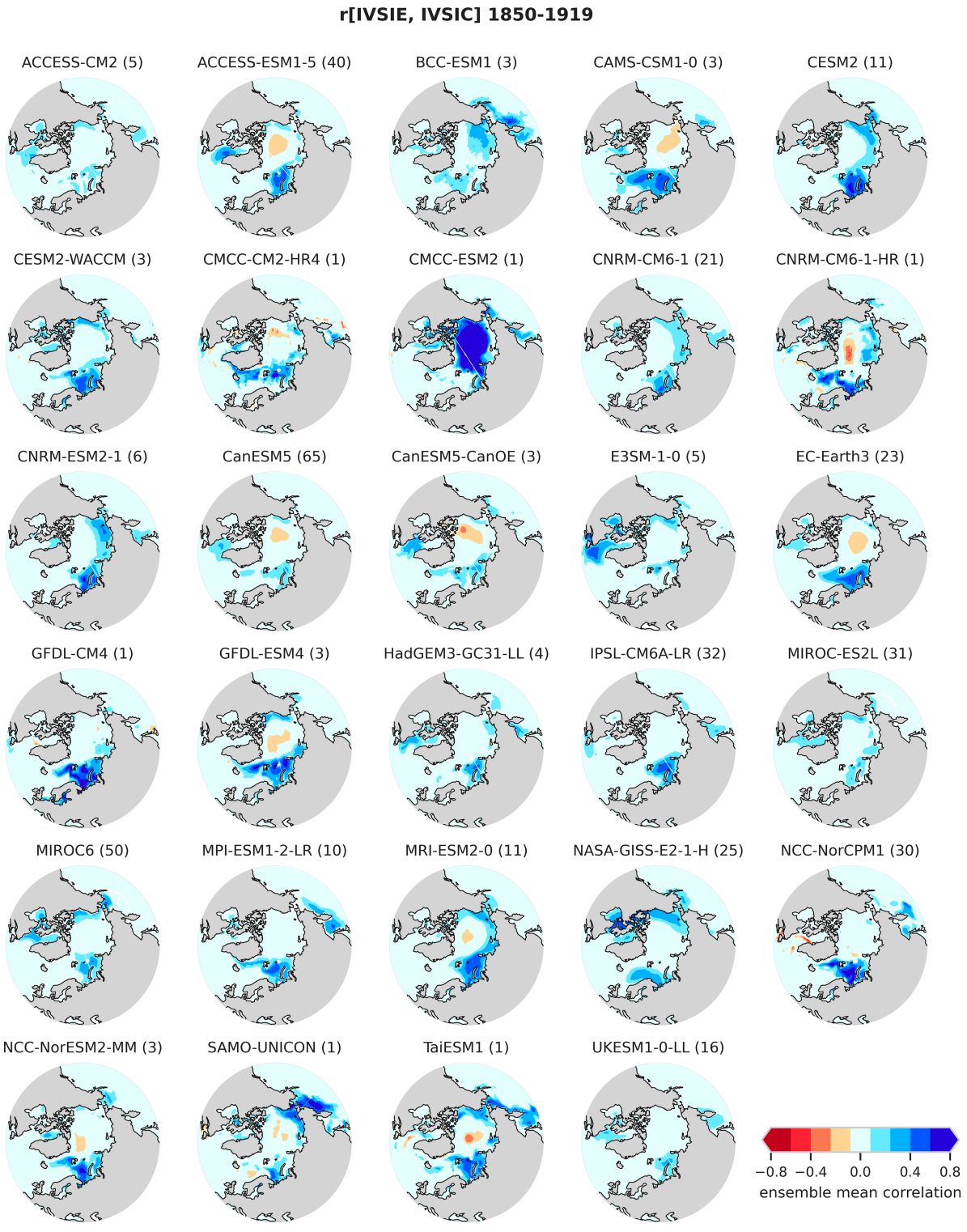

FIG. 4. As in Fig. 3, but for interannual SIE and SIC.

(blue line in Fig. 7b) reveals a pronounced warming during the ETCW (1915–45), followed by a cooling phase (1946–78) and the modern anthropogenic warming (1979 to present). Similar multidecadal variability is observed in global temperatures (black line in Fig. 7b), though with a smaller amplitude.

Figure 7c illustrates simulated trends of annual pan-Arctic SIE anomalies in the 29 CMIP6 models over the 30-yr period

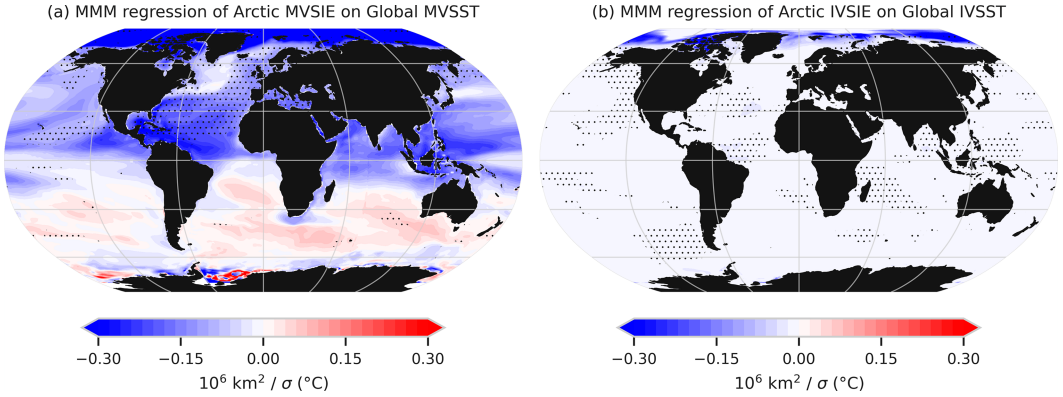

FIG. 5. MMM regressions of Arctic SIE on standardized global SST for (a) multidecadal and (b) interannual anomalies. Only models with at least 10 members are included in the multimodel ensemble. Stippling denotes 95% significance according to a Student's *t* test.

corresponding to the ETCW (1915–45) together with observational estimates using the B20SIE reconstructions. Ensemble mean trends (marked in diamonds) range from −0.31 to 0.15 million km^2 decade^{-1}, with some models showing relatively pronounced sea ice decline (e.g., CMCC-ESM2, GFDL-CM4, and NASA-GISS-E2-1-H) and others exhibiting little or no sea ice loss (e.g., GFDL-ESM4, HadGEM3-GC31-LL, and MPI-ESM1-2-LR). With exception for CNRM-ESM2-1 and NCC-NorCPM1, which show positive ensemble mean SIE trends during the ETCW, all CMIP6 models have ensemble mean trends below zero. However, when considering models with at least 10 members, negative trends are only statistically significant in less than half of them (10 of out 23; bold diamonds in Fig. 7c).

Furthermore, only about two thirds of models have at least one ensemble member that falls within one standard deviation of the B20SIE mean trend, and the CMIP6 MMM SIE trend during the ETCW is −0.12 million km^2 decade^{-1}, significantly smaller than the B20SIE estimates of −0.24/−0.34 million km^2 decade^{-1}. This illustrates that while CMIP6 models tend to simulate a decline in SIE during the ETCW, they generally underestimate the extent of sea ice loss compared to sea ice reconstructions.

The ensemble spread of trends within individual models illustrates the role that internal climate variability may have had on the observed sea ice decline during the ETCW. When normalizing ETCW SIE trends relative to the 1850–1919

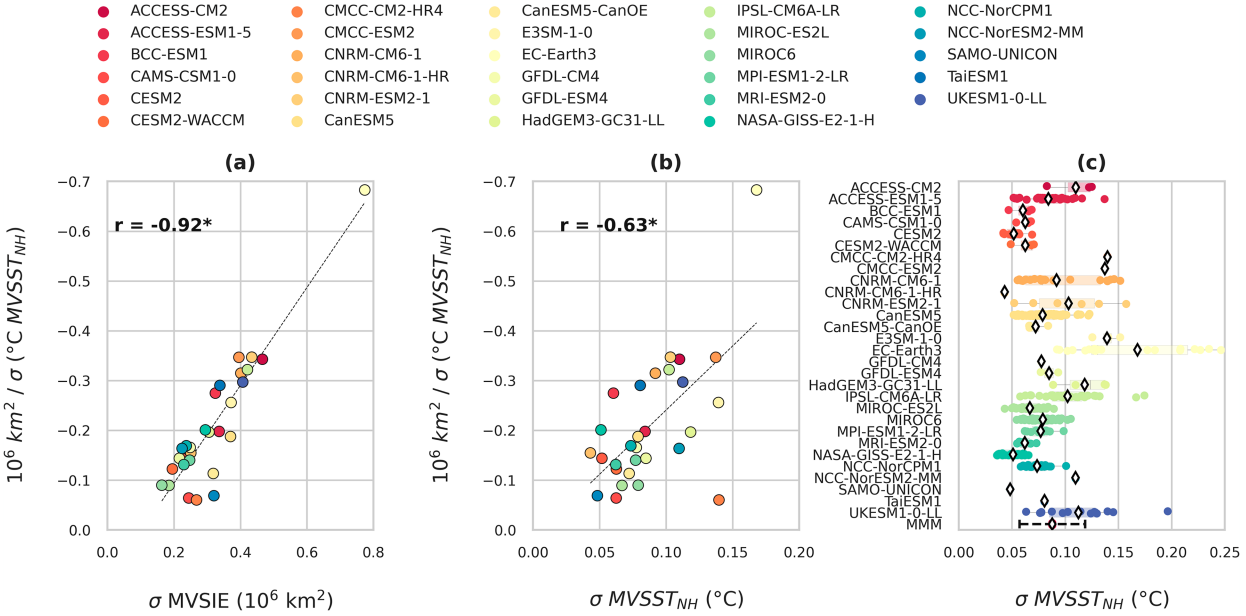

FIG. 6. (a) Correlation between each model's ensemble mean multidecadal SIE variability (σ_{MVSIE}) and the regression of MVSIE onto standardized SST in the Northern Hemisphere ($\mathrm{MVSST_{NH}}$). (b) As in (a), but showing the correlation between ensemble mean $\mathrm{MVSST_{NH}}$ variability ($\sigma_{\mathrm{MVSST_{NH}}}$) and the regression of MVSIE onto standardized $\mathrm{MVSST_{NH}}$. (c) The $\sigma_{\mathrm{MVSST_{NH}}}$ expressed as standard deviations between 1850 and 1919 in the 29 CMIP6 models, with individual members shown with colored dots and ensemble means marked in diamonds.

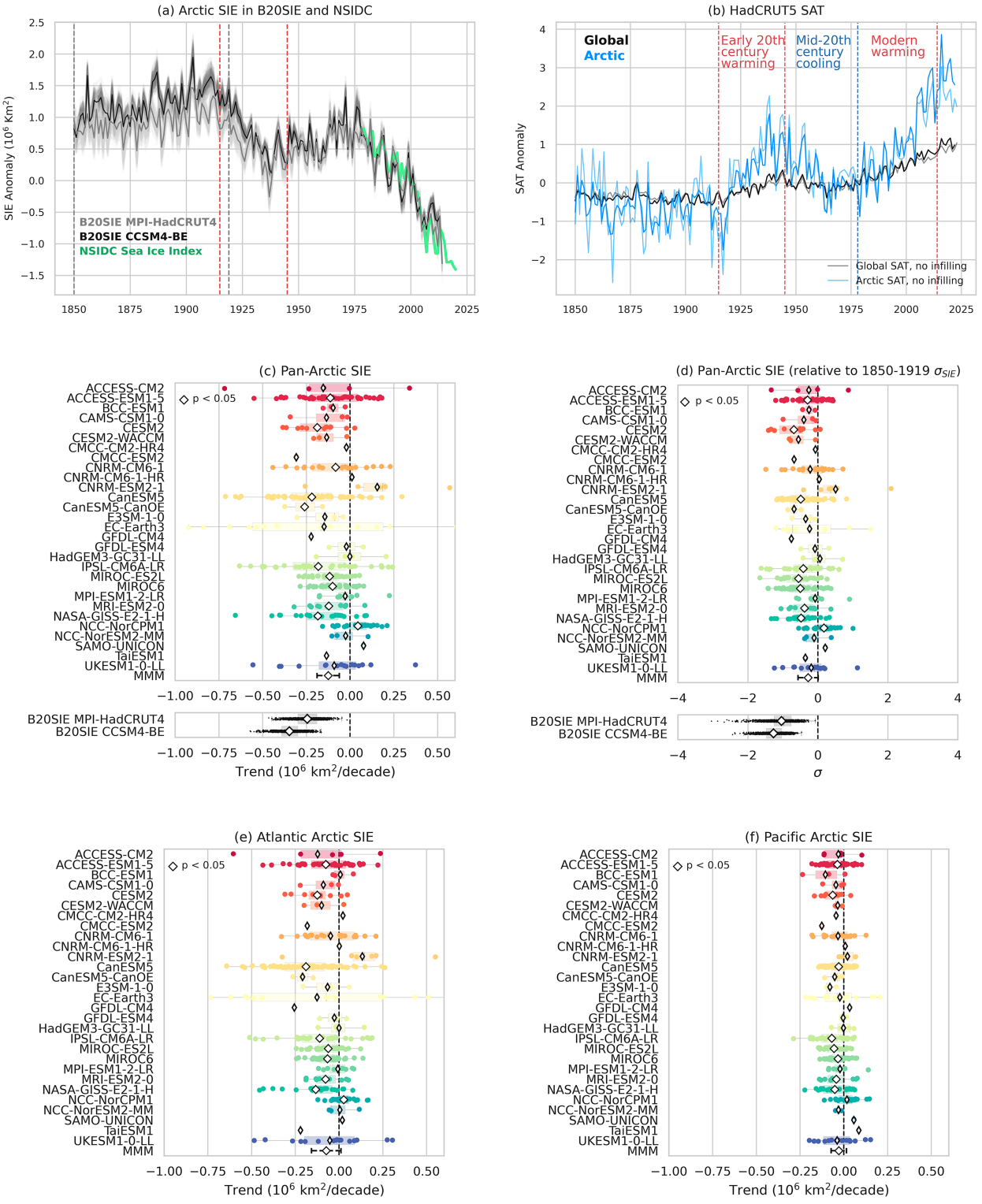

FIG. 7. (a) Arctic SIE anomalies in B20SIE reconstructions and NSIDC observations. Bold lines mark the ensemble mean of B20SIE. Anomalies are centered about 1979–2013. Gray dashed lines mark the 1850–1919 period, and red dashed lines indicate the period corresponding to the ETCW. (b) Global (black) and Arctic (blue) temperature observations from HadCRUT5 analysis, expressed as anomalies relative to 1961–90. Thinner lines indicate the version of HadCRUT5 with no statistical infilling. (c) Linear trends of Arctic SIE during the ETCW in 29 CMIP6 models and B20SIE reconstructions. Diamonds mark ensemble mean SIE trends. Bold diamonds denote statistically significant ensemble mean trends. (d) As in (b), except that ETCW SIE trends are expressed relative to the models' background SIE variability in 1850–1919. (e),(f) As in (c), except that trends are calculated separately for the (e) Atlantic Arctic and (f) the Pacific Arctic.

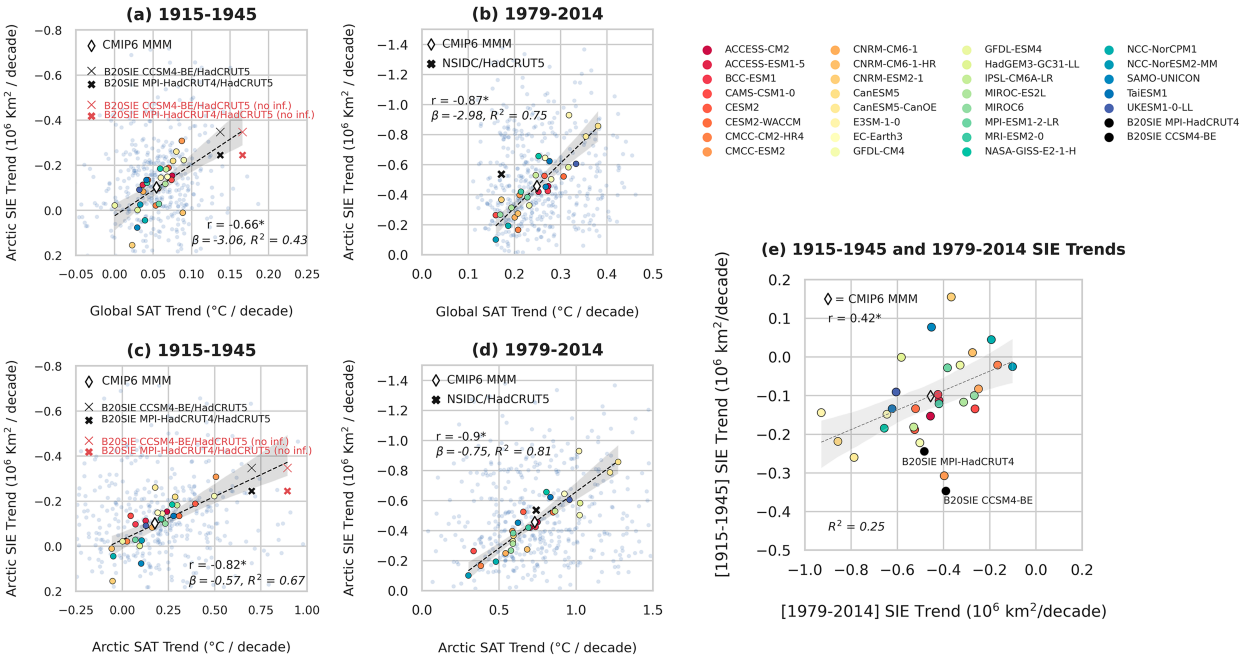

FIG. 8. Arctic SIE trends vs (a),(b) global SAT and (c),(d) Arctic (70°–90°N) SAT trends for (a),(c) the ETCW (1915–45) and (b),(d) the recent warming period (1979–2014). Ensemble mean trends of CMIP6 models are marked in color, while pale blue dots represent individual members. For comparison, observed trends (HadCRUT5 and B20SIE for the ETCW; HadCRUT5 and NSIDC for the modern period) are marked with an "x." For the ETCW, we also show SAT trends for the version of HadCRUT5 with no statistical infilling (red "x"). (e) Ensemble mean SIE trends of CMIP6 models and B20SIE reconstructions in 1915–45 and 1979–2014. Dashed lines mark the linear regression fit with shading indicating 95% confidence intervals based on bootstrapping. Correlation and regression coefficients in all panels refer exclusively to CMIP6 ensemble means.

background SIE variability [i.e., as the ratio of ETCW SIE trends to $\sigma_{\mathrm{SIE}(1850–1919)}$, Fig. 7d], none of the models replicate the reconstructed ETCW sea ice loss ($<-1\sigma$), even if some CMIP6 ensemble mean trends approach the reconstructed mean trend shown in Fig. 7c. This suggests that either the B20SIE reconstructions have too low a SIE variability during 1850–1919, or that CMIP6 models fail to capture the full magnitude of the ETCW (if the ETCW was mostly forced) and/or that the ETCW was an extreme realization of internal variability.

Moreover, partitioning of ETCW SIE trends between the Atlantic- and Pacific-influenced sectors of the Arctic Ocean (defined by dividing the Arctic region longitudinally along the 90° and 270° meridians) reveals that the SIE decline is almost entirely limited to the Atlantic Arctic (Figs. 7e,f). This finding is consistent with previous observations of temperature anomalies during the ETCW, which also showed a similar spatial distribution dominated by warming in the Atlantic-adjacent Arctic (Johannessen et al. 2004).

To explore the sources of model spread in ETCW SIE trends, we analyze Arctic sea ice trends in relation to global and Arctic SAT. Figure 8 shows scatterplots of Arctic SIE versus global and Arctic SAT trends during the ETCW (1915–45) and in the recent warming period (1979–2014) in CMIP6 and observations. As expected from Fig. 5 and previous literature, models with greater global and Arctic temperature trends show greater SIE trends. The correlation between Arctic SIE and global SAT trends in CMIP6 model ensemble means is stronger during 1979–2014 ($r = -0.87$) than during the ETCW ($r = -0.66$; Figs. 8a,b). However, the simulated sensitivity of SIE loss to an increase in global SAT is consistent across the two periods, as indicated by regression coefficients of -2.98 million and -3.06 million km^2 °C^{-1} during 1979–2014 and 1915–45, respectively. During the ETCW, the simulated sensitivity is generally consistent with the corresponding observed sensitivity (e.g., $-2.53/-2.09$ million km^2 °C^{-1} for B20SIE-CCSM4/HadCRUT5). This suggests that the biased representation of sea ice loss in CMIP6 models stems from not simulating sufficient warming during the ETCW, rather than from a biased sensitivity to the temperature trend.

The correlation between Arctic SIE and Arctic SAT trends (Figs. 8c,d) is consistently strong across the two periods ($r = -0.9$ and $r = -0.82$ in 1979–2014 and 1915–45, respectively), supporting previous evidence that the ETCW temperature anomaly was primarily characterized by warming at high northern latitudes (e.g., Johannessen et al. 2004). The sensitivity of SIE loss to Arctic warming is higher during the modern period than during the ETCW, both in models (-0.75 vs -0.57 million km^2 °C^{-1}) and observations (-0.87 for NSIDC/HadCRUT5 vs -0.49 million km^2 °C^{-1} for B20SIE-CCSM4/HadCRUT5, in 1979–2014 and 1915–45, respectively). It is worth noting that changes in observed sensitivity to Arctic and global temperatures across the two periods could be in part explained by internal variability, which over recent decades has contributed to enhanced Arctic warming but dampened global warming (Sweeney et al. 2023).

We additionally find a significant correlation between CMIP6 ensemble mean trends of Arctic SIE in the two periods ($r = 0.42$; Fig. 8e), suggesting that models that simulate weaker SIE trends during the ETCW generally tend to underestimate the recent sea ice decline, and vice versa. Models with lower sensitivity to global and Arctic SAT, such as CMCC-CM2-HR4, NCC-NorCPM1, and NCC-NorESM2-MM (Figs. 8a–d), underestimate ensemble mean SIE trends in both periods relative to the CMIP6 MMM and B20SIE reconstructions (Fig. 8e). Conversely, CanESM5, CanESM5-CanOE, and E3SM-1-0, which exhibit higher sensitivity to SAT, are better able to capture the ETCW but overestimate the 1979–2014 sea ice decline.

To further assess the time evolution of multidecadal changes in CMIP6 and observations, we calculate linear trends of Arctic SIE, Arctic SAT, and global SAT for successive segments of approximately 30 years (1915–45, 1946–78, and 1979–2014). Figure 9 shows the multimodel distribution of CMIP6 trends along with B20SIE reconstructions and observations over the three subperiods. Similarly to CMIP5 models, which showed reduced Arctic warming during the ETCW compared to observations (e.g., Fyfe et al. 2013), CMIP6 models, as a group, underestimate the magnitude of the ETCW event with respect to temperature observations, consistent with the weaker SIE trends relative to the B20SIE reconstructions. It should be noted that confidence in the accuracy of observations during the instrumental era is reduced due to sparse and heterogeneous data acquisition (e.g., Morice et al. 2021), as further indicated by the discrepancy between observed trends in the versions with and without infilling of HadCRUT5 (Figs. 8 and 9). However, given the multimodel ensemble of 360+ realizations, which can be regarded as a representative sample of internal climate variability trajectories, our results show that CMIP6 models are deficient in their representation of the ETCW. Nevertheless, models and observations show better agreement in sea ice and temperature changes during the mid-twentieth century cooling period (1946–78) and the modern anthropogenic warming (1979–2014). The 1946–78 observed Arctic cooling is slightly underestimated by models, in agreement with England et al. (2021). The spread of individual model realizations is pronounced in all subperiods, suggesting a critical contribution of internal climate variability to multidecadal changes. We note that the CMIP6 multimodel trends depicted in Fig. 9 are not impacted by the different ensemble sizes of models, and the distributions do not change when limiting the number of ensemble members per model to $n = 20$ (not shown).

4. Discussion and conclusions

The current understanding of Arctic sea ice variability is limited by the relatively short temporal coverage of satellite observations and by the uncertainty in GCM simulations that arises from model bias and background climate variability. Assessing model performance in historical simulations is a critical task, as the robustness of future projections hinges on the demonstrated ability of models to realistically represent changes in the past climate.

Here, we have analyzed Arctic sea ice variability on interannual and multidecadal time scales using historical simulations from a subset of 29 CMIP6 models with all available realizations. Analysis of the portion of the historical time series with low external forcing (1850–1919) reveals that models display realistic interannual sea ice variability relative to sea ice reconstructions (Fig. 1b), with relatively little intermodel spread (Fig. 2d). This finding is consistent with Wyburn-Powell et al. (2022), who demonstrated good agreement between CMIP5 single model initial-condition large ensemble (SMILE) models and observations in terms of interannual sea ice variability, albeit with seasonal and regional differences. On the other hand, we found substantial uncertainty in models' representation of multidecadal sea ice variability, both in terms of intermodel spread (Figs. 1a and 2c) and differences with respect to the B20SIE sea ice reconstructions (Fig. 1a). The subpolar North Atlantic is the key region determining model uncertainty in MVSIE, likely tied to ocean forcing of sea ice (e.g., Bitz et al. 2005), and model uncertainty in the coupling of ocean modes of variability (such as the AMOC) and North Atlantic SST (e.g., Tandon and Kushner 2015; Liu and Fedorov 2022). Our results highlight different sources of intermodel spread in sea ice variability, which differ across time scales. The intermodel spread of interannual sea ice variability is in part explained by mean-state biases (Fig. 1d). This finding has emerged in previous studies (e.g., Shu et al. 2020), though with considerable seasonal variations (Blanchard-Wrigglesworth et al. 2021). On multidecadal time scales, sensitivity to Northern Hemispheric temperatures contributes substantially to the intermodel spread of SIE variability (Figs. 5 and 6). Models with biased low sensitivity to Northern Hemispheric SST tend to underestimate MVSIE variability with respect to the CMIP6 multimodel mean, particularly in the Greenland and Labrador Seas (Fig. 3; Figs. S3 and S4), and vice versa. Interestingly, the simulated NH SST variability itself is correlated to the sensitivity of SIE to SST (Fig. 6b), with models that simulate greater SST variability also simulating greater SIE to SST sensitivity (even after normalizing SST variability). This result supports the notion that a realistic representation of long-term sea ice changes depends on models' ability to accurately simulate sea ice sensitivity to large-scale climate variability, as pointed out in previous studies (Mahlstein and Knutti 2012; Rosenblum and Eisenman 2017; Notz and SIMIP Community 2020; Long et al. 2021). However, it is worth noting that estimates of simulated sea ice sensitivity cannot be adequately interpreted without taking into account the influence of internal climate variability (Ding et al. 2019), meaning that these two aspects should not be regarded as separate sources of uncertainty among GCMs.

It remains unclear which underlying physical processes are represented differently in models with sensitivity to Northern Hemispheric SST (or SST variability) that depart substantially from the CMIP6 multimodel mean. While it is challenging to determine direct causal links to specific model deficiencies, the large intermodel spread on multidecadal time scales (reflected both on Arctic SIE and Northern Hemispheric SST; Figs. 1 and 6) is consistent with widely documented uncertainties in the representation of ocean circulation and dynamics

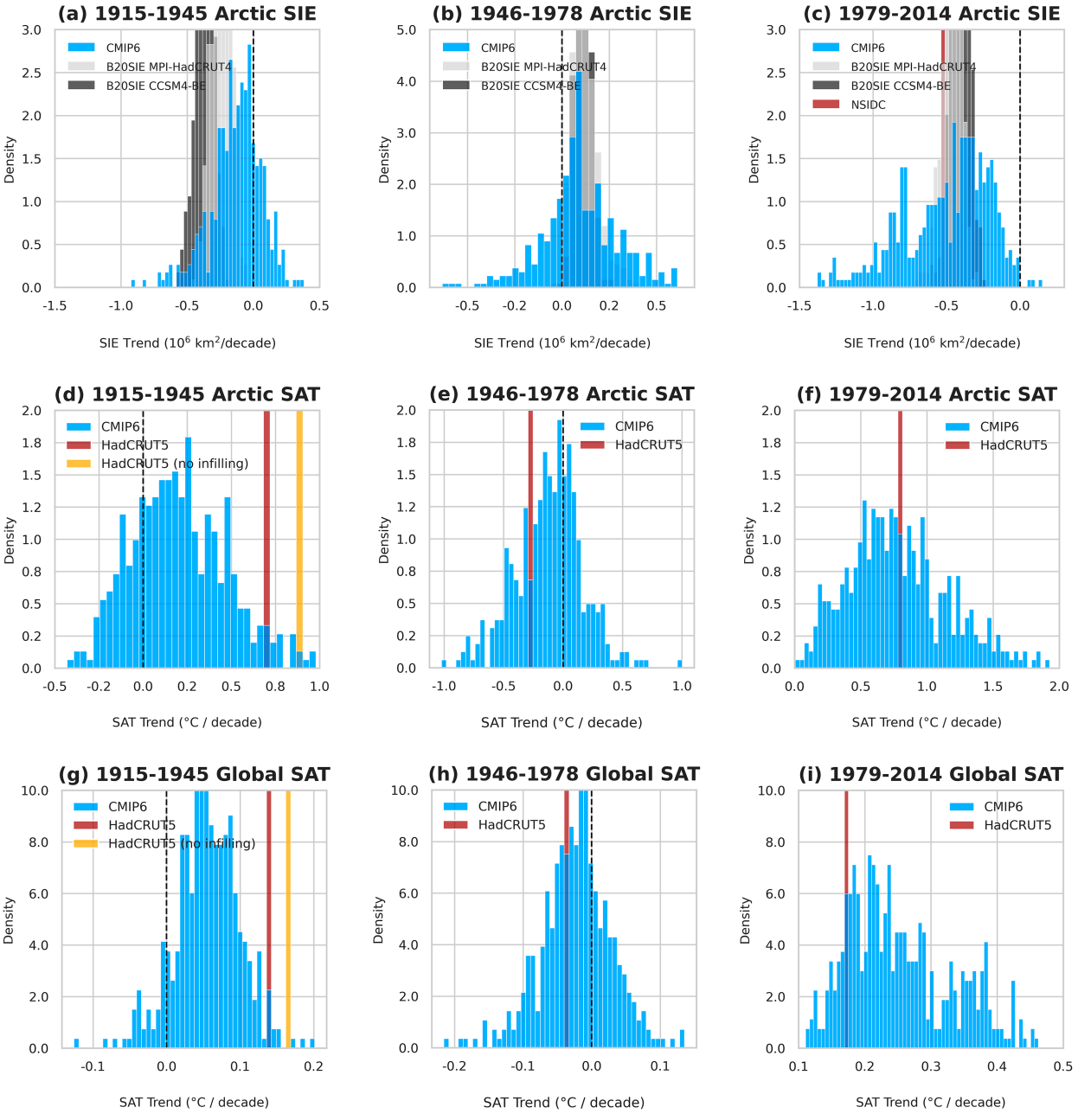

FIG. 9. Distribution of linear trends of (a)–(c) Arctic SIE, (d)–(f) Arctic SAT, and (g)–(i) global SAT for the CMIP6 multimodel ensemble (366 members from 29 models), B20SIE reconstructions (200 members, two model priors), and observations (NSIDC for SIE, HadCRUT5 for SAT) for three segments of the historical time series: (a),(d),(g) the ETCW (1915–45), (b),(e),(h) the intermediate cooling period following the ETCW (1946–78), and (c),(f),(i) the modern period (1978–2014). The *y* axes refer to the normalized density of members falling into each bin, with dashed lines highlighting the *y* axis at 0.

within CMIP6 (e.g., Pan et al. 2023; Cai et al. 2021). For instance, Winkelbauer et al. (2024) showed that intermodel spread in simulated ocean heat transports, mostly due to temperature biases, is strongly associated with a biased representation of the mean Arctic state within CMIP6. Other processes that are believed to be related to the AMOC, a dominant source of multidecadal variability in the Arctic, still suffer from inadequate representation and intermodel uncertainty (Fox-Kemper et al. 2019), including mixed-layer depth (e.g., Jackson and Petit 2023), deep water properties (Heuzé et al. 2023), sea surface salinity, and liquid freshwater content (Wang et al. 2022; Zanowski et al. 2021). Model discrepancies in upper ocean salinity and stratification in the Arctic (Muilwijk et al. 2023; Khosravi et al. 2022), which regulate the influence of the

Atlantic water layer on sea ice, are a likely source of the pronounced model spread of MVSIE variability in the subpolar North Atlantic (Fig. 2c). However, more in-depth studies are needed to quantify these deficiencies, in order to identify specific areas for improvement in the realism of Arctic climate variability for the next generations of GCMs.

Moreover, we have shown that while CMIP6 models are generally capable of reproducing multidecadal sea ice changes during the recent warming period (1979–2014) and the preceding cooling period (1946–78) relative to observations (Figs. 9b,c), they tend to underestimate the sea ice decline associated with the ETCW (Figs. 7c and 9a). Simulated trends appear especially small when examined in relation to the background sea ice variability (1850–1919; Fig. 7d), which is greater in CMIP6 than in observations (Fig. 1a). This aspect raises an important question. If CMIP6 models overestimate the pre-ETCW multidecadal variability, then why do over a third of them fail to simulate the ETCW sea ice decline? There are two possible explanations for this discrepancy. The first is that there is insufficient internal variability in the sea ice reconstructions. The B20SIE product has demonstrated a substantial improvement from previous reconstructions of Arctic sea ice prior to the satellite era (Walsh et al. 2017), both in terms of larger sea ice variability and in terms of greater sea ice loss and subsequent recovery during and after the ETCW, respectively (B20SIE; Brennan and Hakim 2022). Nevertheless, it is possible that multidecadal variability before 1920 is not fully captured in the B20SIE product due to biases inherited from the selected model priors (MPI-ESM-P and CCSM4) and uncertainties in early twentieth century observations. Moreover, the annual frequency of B20SIE reconstructions prevents the assessment of seasonal changes in Arctic sea ice during the instrumental era.

The second alternative explanation is that the ETCW was at least in part an externally forced event and models do not simulate a correct response to early twentieth century forcing. Given the relatively good agreement between simulated and observed sensitivities to the ETCW (Figs. 8a,c), it is apparent that the biased low SIE loss during this period is largely due to the fact that CMIP6 models, as a group, simulate insufficient warming of both Arctic and global SAT (Figs. 8 and 9). This finding is in line with previous assessments by Papalexiou et al. (2020) and Latonin et al. (2021). Fyfe et al. (2013) additionally concluded that the deficiency of CMIP5 models in capturing the observed Arctic warming during the ETCW was linked to an incomplete representation of sulfate and black carbon aerosols. Recent studies have shown that aerosol-related forcing can have large effects on multidecadal variations in Arctic sea ice, but the spread in the representation of aerosol emissions within CMIP6 remains large (DeRepentigny et al. 2022; Aizawa et al. 2022). Additionally, natural forcings have been shown to play a significant role in the simulated magnitude of the Arctic SAT signal during the ETCW (Aizawa et al. 2021). While our study did not specifically investigate the relative contributions of external forcings and internal climate variability to the ETCW, the pronounced difference in temperature trends between models and observations, given the large number of realizations considered, points to an insufficient representation of the forced response.

To reduce the uncertainty in Arctic sea ice projections, more research work should focus on improving our understanding of multidecadal sea ice variability and its drivers. Since models show relatively good agreement in their representation of interannual sea ice variability, which is mostly driven by Arctic processes, what determines the credibility of a model is its ability to simulate the sea ice response to low-frequency hemispheric and global climate variability. More crucially, the role of internal climate variability in model projections may be inadequately represented if models overestimate low-frequency sea ice variability in the past. The model biases highlighted in this study mainly originate from the representation of processes coupling Arctic sea ice to Northern Hemispheric climate and encompassing different components of the climate system. Thus, these processes and their interactions should be prioritized in order to improve the realism of next-generation climate models.

Acknowledgments. EB was supported by the PhD program in Science and Management of Climate Change at Ca' Foscari University. EB-W was supported by NSF Grant OPP-2213988. DI was supported by the European Union's Horizon 2020 research and innovation program under Grant Agreement 615 No. 101003826 via the project CRiceS. SM was supported by the Foundation Euro-Mediterranean Center on Climate Change (CMCC, Italy). This research was supported by the European Union Horizon 2020 research and innovation programme under Grant Agreement 101003826 via the project CRiceS (Climate-relevant interactions and feedbacks: the key role of sea ice and snow in the polar and global climate system). We would also like to acknowledge high-performance computing support from Cheyenne (https://doi.org/10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation.

Data availability statement. CMIP6 data are made available by the ESGF (https://esgf-node.llnl.gov/search/cmip6/) and can be downloaded from various ESGF nodes and portals (e.g., https://aims2.llnl.gov/). Arctic SIE reconstructions by B20SIE are available online at https://zenodo.org/record/3717240. The NSIDC Sea Ice Index, version 3, can be accessed at https://nsidc.org/data/g02135/versions/3. HadCRUT5 temperature observations are provided by the Met Office Hadley Center at https://crudata.uea.ac.uk/cru/data/temperature/.

REFERENCES

Aizawa, T., M. Ishii, N. Oshima, S. Yukimoto, and H. Hasumi, 2021: Arctic warming and associated sea ice reduction in the early 20th century induced by natural forcings in MRI-ESM2.0 climate simulations and multimodel analyses. *Geophys. Res. Lett.*, **48**, e2020GL092336, https://doi.org/10.1029/2020GL092336.

——, N. Oshima, and S. Yukimoto, 2022: Contributions of anthropogenic aerosol forcing and multidecadal internal variability

to mid-20th century Arctic cooling—CMIP6/DAMIP multimodel analysis. *Geophys. Res. Lett.*, **49**, e2021GL097093, https://doi.org/10.1029/2021GL097093.

Baxter, I., and Coauthors, 2019: How tropical Pacific surface cooling contributed to accelerated sea ice melt from 2007 to 2012 as ice is thinned by anthropogenic forcing. *J. Climate*, **32**, 8583–8602, https://doi.org/10.1175/JCLI-D-18-0783.1.

Beitsch, A., J. H. Jungclaus, and D. Zanchettin, 2014: Patterns of decadal-scale Arctic warming events in simulated climate. *Climate Dyn.*, **43**, 1773–1789, https://doi.org/10.1007/s00382-013-2004-5.

Bengtsson, L., V. A. Semenov, and O. M. Johannessen, 2004: The early twentieth-century warming in the Arctic—A possible mechanism. *J. Climate*, **17**, 4045–4057, https://doi.org/10.1175/1520-0442(2004)017<4045:TETWIT>2.0.CO;2.

Bitz, C. M., D. S. Battisti, R. E. Moritz, and J. A. Beesley, 1996: Low-frequency variability in the Arctic atmosphere, sea ice, and upper-ocean climate system. *J. Climate*, **9**, 394–408, https://doi.org/10.1175/1520-0442(1996)009<0394:LFVITA>2.0.CO;2.

——, M. M. Holland, E. C. Hunke, and R. E. Moritz, 2005: Maintenance of the sea-ice edge. *J. Climate*, **18**, 2903–2921, https://doi.org/10.1175/JCLI3428.1.

Blanchard-Wrigglesworth, E., A. Donohoe, L. A. Roach, A. DuVivier, and C. M. Bitz, 2021: High-frequency sea ice variability in observations and models. *Geophys. Res. Lett.*, **48**, e2020GL092356, https://doi.org/10.1029/2020GL092356.

Bokuchava, D. D., and V. A. Semenov, 2021: Mechanisms of the early 20th century warming in the Arctic. *Earth-Sci. Rev.*, **222**, 103820, https://doi.org/10.1016/j.earscirev.2021.103820.

Bonan, D. B., and E. Blanchard-Wrigglesworth, 2020: Nonstationary teleconnection between the Pacific Ocean and Arctic sea ice. *Geophys. Res. Lett.*, **47**, e2019GL085666, https://doi.org/10.1029/2019GL085666.

Brennan, M. K., and G. J. Hakim, 2022: Reconstructing Arctic sea ice over the Common Era using data assimilation. *J. Climate*, **35**, 1231–1247, https://doi.org/10.1175/JCLI-D-21-0099.1.

——, ——, and E. Blanchard-Wrigglesworth, 2020: Arctic sea-ice variability during the Instrumental Era. *Geophys. Res. Lett.*, **47**, e2019GL086843, https://doi.org/10.1029/2019GL086843.

Cai, Z., Q. You, F. Wu, H. W. Chen, D. Chen, and J. Cohen, 2021: Arctic warming revealed by multiple CMIP6 models: Evaluation of historical simulations and quantification of future projection uncertainties. *J. Climate*, **34**, 4871–4892, https://doi.org/10.1175/JCLI-D-20-0791.1.

Cavalieri, D. J., C. L. Parkinson, and K. Y. Vinnikov, 2003: 30-year satellite record reveals contrasting Arctic and Antarctic decadal sea ice variability. *Geophys. Res. Lett.*, **30**, 1970, https://doi.org/10.1029/2003GL018031.

Clancy, R., C. Bitz, and E. Blanchard-Wrigglesworth, 2021: The influence of ENSO on Arctic sea ice in large ensembles and observations. *J. Climate*, **34**, 9585–9604, https://doi.org/10.1175/JCLI-D-20-0958.1.

Comiso, J. C., C. L. Parkinson, R. Gersten, and L. Stock, 2008: Accelerated decline in the Arctic sea ice cover. *Geophys. Res. Lett.*, **35**, L01703, https://doi.org/10.1029/2007GL031972.

Davy, R., and S. Outten, 2020: The Arctic surface climate in CMIP6: Status and developments since CMIP5. *J. Climate*, **33**, 8047–8068, https://doi.org/10.1175/JCLI-D-19-0990.1.

DeRepentigny, P., and Coauthors, 2022: Enhanced simulated early 21st century Arctic sea ice loss due to CMIP6 biomass burning emissions. *Sci. Adv.*, **8**, eabo2405, https://doi.org/10.1126/sciadv.abo2405.

Ding, Q., and Coauthors, 2017: Influence of high-latitude atmospheric circulation changes on summertime Arctic sea ice. *Nat. Climate Change*, **7**, 289–295, https://doi.org/10.1038/nclimate3241.

——, and Coauthors, 2019: Fingerprints of internal drivers of Arctic sea ice loss in observations and model simulations. *Nat. Geosci.*, **12**, 28–33, https://doi.org/10.1038/s41561-018-0256-8.

Dörr, J. S., D. B. Bonan, M. Årthun, L. Svendsen, and R. C. J. Wills, 2023: Forced and internal components of observed Arctic sea-ice changes. *Cryosphere*, **17**, 4133–4153, https://doi.org/10.5194/tc-17-4133-2023.

England, M., A. Jahn, and L. Polvani, 2019: Nonuniform contribution of internal variability to recent Arctic sea ice loss. *J. Climate*, **32**, 4039–4053, https://doi.org/10.1175/JCLI-D-18-0864.1.

England, M. R., I. Eisenman, N. J. Lutsko, and T. J. W. Wagner, 2021: The recent emergence of Arctic amplification. *Geophys. Res. Lett.*, **48**, e2021GL094086, https://doi.org/10.1029/2021GL094086.

Fetterer, F., K. Knowles, W. N. Meier, M. Savoie, and A. K. Windnagel, 2017: Sea ice index, version 3. National Snow and Ice Data Center, accessed 18 September 2024, https://doi.org/10.7265/N5K072F8.

Fox-Kemper, B., and Coauthors, 2019: Challenges and prospects in ocean circulation models. *Front. Mar. Sci.*, **6**, 65, https://doi.org/10.3389/fmars.2019.00065.

Fyfe, J. C., K. v. Salzen, N. P. Gillett, V. K. Arora, G. M. Flato, and J. R. McConnell, 2013: One hundred years of Arctic surface temperature variation due to anthropogenic influence. *Sci. Rep.*, **3**, 2645, https://doi.org/10.1038/srep02645.

Gregory, J. M., P. A. Stott, D. J. Cresswell, N. A. Rayner, C. Gordon, and D. M. H. Sexton, 2002: Recent and future changes in Arctic sea ice simulated by the HadCM3 AOGCM. *Geophys. Res. Lett.*, **29**, 2175, https://doi.org/10.1029/2001GL014575.

Hegerl, G. C., S. Brönnimann, A. Schurer, and T. Cowan, 2018: The early 20th century warming: Anomalies, causes, and consequences. *Wiley Interdiscip. Rev.: Climate Change*, **9**, e522, https://doi.org/10.1002/wcc.522.

Heuzé, C., H. Zanowski, S. Karam, and M. Muilwijk, 2023: The deep Arctic Ocean and Fram Strait in CMIP6 models. *J. Climate*, **36**, 2551–2584, https://doi.org/10.1175/JCLI-D-22-0194.1.

Jackson, L. C., and T. Petit, 2023: North Atlantic overturning and water mass transformation in CMIP6 models. *Climate Dyn.*, **60**, 2871–2891, https://doi.org/10.1007/s00382-022-06448-1.

Jahn, A., M. M. Holland, and J. E. Kay, 2024: Projections of an ice-free Arctic Ocean. *Nat. Rev. Earth Environ.*, **5**, 164–176, https://doi.org/10.1038/s43017-023-00515-9.

Johannessen, O. M., and Coauthors, 2004: Arctic climate change: Observed and modelled temperature and sea-ice variability. *Tellus*, **56A**, 328–341, https://doi.org/10.3402/tellusa.v56i4.14418.

Kay, J. E., M. M. Holland, and A. Jahn, 2011: Inter-annual to multi-decadal Arctic sea ice extent trends in a warming world. *Geophys. Res. Lett.*, **38**, L15708, https://doi.org/10.1029/2011GL048008.

Khosravi, N., Q. Wang, N. Koldunov, C. Hinrichs, T. Semmler, S. Danilov, and T. Jung, 2022: The Arctic Ocean in CMIP6 models: Biases and projected changes in temperature and salinity. *Earth's Future*, **10**, e2021EF002282, https://doi.org/10.1029/2021EF002282.

Latonin, M. M., I. L. Bashmachnikov, L. P. Bobylev, and R. Davy, 2021: Multi-model ensemble mean of global climate models fails to reproduce early twentieth century Arctic

warming. *Polar Sci.*, **30**, 100677, https://doi.org/10.1016/j.polar.2021.100677.

Li, D., R. Zhang, and T. R. Knutson, 2017: On the discrepancy between observed and CMIP5 multi-model simulated Barents Sea winter sea ice decline. *Nat. Commun.*, **8**, 14991, https://doi.org/10.1038/ncomms14991.

Liu, W., and A. Fedorov, 2022: Interaction between Arctic sea ice and the Atlantic meridional overturning circulation in a warming climate. *Climate Dyn.*, **58**, 1811–1827, https://doi.org/10.1007/s00382-021-05993-5.

Long, M., L. Zhang, S. Hu, and S. Qian, 2021: Multi-aspect assessment of CMIP6 models for Arctic sea ice simulation. *J. Climate*, **34**, 1515–1529, https://doi.org/10.1175/JCLI-D-20-0522.1.

Mahlstein, I., and R. Knutti, 2012: September Arctic sea ice predicted to disappear near 2°C global warming above present. *J. Geophys. Res.*, **117**, D016104, https://doi.org/10.1029/2011JD016709.

Massonnet, F., M. Vancoppenolle, H. Goosse, D. Docquier, T. Fichefet, and E. Blanchard-Wrigglesworth, 2018: Arctic sea-ice change tied to its mean state through thermodynamic processes. *Nat. Climate Change*, **8**, 599–603, https://doi.org/10.1038/s41558-018-0204-z.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadCRUT4 data set. *J. Geophys. Res.*, **117**, D08101, https://doi.org/10.1029/2011JD017187.

——, and Coauthors, 2021: An updated assessment of near-surface temperature change from 1850: The hadCRUT5 data set. *J. Geophys. Res. Atmos.*, **126**, e2019JD032361, https://doi.org/10.1029/2019JD032361.

Muilwijk, M., A. Nummelin, L. H. Smedsrud, I. V. Polyakov, H. Zanowski, and C. Heuzé, 2023: Divergence in climate model projections of Arctic atlantification. *J. Climate*, **36**, 1727–1748, https://doi.org/10.1175/JCLI-D-22-0349.1.

Notz, D., and SIMIP Community, 2020: Arctic sea ice in CMIP6. *Geophys. Res. Lett.*, **47**, e2019GL086749, https://doi.org/10.1029/2019GL086749.

Olonscheck, D., T. Mauritsen, and D. Notz, 2019: Arctic sea-ice variability is primarily driven by atmospheric temperature fluctuations. *Nat. Geosci.*, **12**, 430–434, https://doi.org/10.1038/s41561-019-0363-1.

Overland, J. E., and M. Wang, 2005: The third Arctic climate pattern: 1930s and early 2000s. *Geophys. Res. Lett.*, **32**, L23808, https://doi.org/10.1029/2005GL024254.

Pan, R., Q. Shu, Q. Wang, S. Wang, Z. Song, Y. He, and F. Qiao, 2023: Future Arctic climate change in CMIP6 strikingly intensified by NEMO-family climate models. *Geophys. Res. Lett.*, **50**, e2022GL102077, https://doi.org/10.1029/2022GL102077.

Papalexiou, S. M., C. R. Rajulapati, M. P. Clark, and F. Lehner, 2020: Robustness of CMIP6 historical global mean temperature simulations: Trends, long-term persistence, autocorrelation, and distributional shape. *Earth's Future*, **8**, e2020EF001667, https://doi.org/10.1029/2020EF001667.

Polyakov, I. V., R. V. Bekryaev, G. V. Alekseev, U. S. Bhatt, R. L. Colony, M. A. Johnson, A. P. Maskshtas, and D. Walsh, 2003: Variability and trends of air temperature and pressure in the maritime Arctic, 1875–2000. *J. Climate*, **16**, 2067–2077, https://doi.org/10.1175/1520-0442(2003)016<2067:VATOAT>2.0.CO;2.

Rosenblum, E., and I. Eisenman, 2017: Sea ice trends in climate models only accurate in runs with biased global warming. *J. Climate*, **30**, 6265–6278, https://doi.org/10.1175/JCLI-D-16-0455.1.

Shen, Z., A. Duan, D. Li, and J. Li, 2021: Assessment and ranking of climate models in Arctic sea ice cover simulation: From CMIP5 to CMIP6. *J. Climate*, **34**, 3609–3627, https://doi.org/10.1175/JCLI-D-20-0294.1.

Shu, Q., Q. Wang, Z. Song, F. Qiao, J. Zhao, M. Chu, and X. Li, 2020: Assessment of sea ice extent in CMIP6 with comparison to observations and CMIP5. *Geophys. Res. Lett.*, **47**, e2020GL087965, https://doi.org/10.1029/2020GL087965.

Sweeney, A. J., Q. Fu, S. Po-Chedley, H. Wang, and M. Wang, 2023: Internal variability increased Arctic amplification during 1980–2022. *Geophys. Res. Lett.*, **50**, e2023GL106060, https://doi.org/10.1029/2023GL106060.

Tandon, N. F., and P. J. Kushner, 2015: Does external forcing interfere with the AMOC's influence on North Atlantic sea surface temperature? *J. Climate*, **28**, 6309–6323, https://doi.org/10.1175/JCLI-D-14-00664.1.

Tian, T., S. Yang, J. L. Høyer, P. Nielsen-Englyst, and S. Singha, 2024: Cooler Arctic surface temperatures simulated by climate models are closer to satellite-based data than the ERA5 reanalysis. *Commun. Earth Environ.*, **5**, 111, https://doi.org/10.1038/s43247-024-01276-z.

Topál, D., Q. Ding, J. Mitchell, I. Baxter, M. Herein, T. Haszpra, R. Luo, and Q. Li, 2020: An internal atmospheric process determining summertime Arctic sea ice melting in the next three decades: Lessons learned from five large ensembles and multiple CMIP5 climate simulations. *J. Climate*, **33**, 7431–7454, https://doi.org/10.1175/JCLI-D-19-0803.1.

Walsh, J. E., F. Fetterer, J. Scott Stewart, and W. L. Chapman, 2017: A database for depicting Arctic sea ice variations back to 1850. *Geogr. Rev.*, **107**, 89–107, https://doi.org/10.1111/j.1931-0846.2016.12195.x.

Wang, M., J. E. Overland, V. Kattsov, J. E. Walsh, X. Zhang, and T. Pavlova, 2007: Intrinsic versus forced variation in coupled climate model simulations over the Arctic during the twentieth century. *J. Climate*, **20**, 1093–1107, https://doi.org/10.1175/JCLI4043.1.

Wang, S., Q. Wang, M. Wang, G. Lohmann, and F. Qiao, 2022: Arctic Ocean freshwater in CMIP6 coupled models. *Earth's Future*, **10**, e2022EF002878, https://doi.org/10.1029/2022EF002878.

Watts, M., W. Maslowski, Y. J. Lee, J. C. Kinney, and R. Osinski, 2021: A spatial evaluation of Arctic sea ice and regional limitations in CMIP6 historical simulations. *J. Climate*, **34**, 6399–6420, https://doi.org/10.1175/JCLI-D-20-0491.1.

Winkelbauer, S., M. Mayer, and L. Haimberger, 2024: Validation of key Arctic energy and water budget components in CMIP6. *Climate Dyn.*, **62**, 3891–3926, https://doi.org/10.1007/s00382-024-07105-5.

Winton, M., 2011: Do climate models underestimate the sensitivity of Northern Hemisphere sea ice cover? *J. Climate*, **24**, 3924–3934, https://doi.org/10.1175/2011JCLI4146.1.

Wood, K. R., and J. E. Overland, 2010: Early 20th century Arctic warming in retrospect. *Int. J. Climatol.*, **30**, 1269–1279, https://doi.org/10.1002/joc.1973.

Wyburn-Powell, C., A. Jahn, and M. R. England, 2022: Modeled interannual variability of Arctic sea ice cover is within observational uncertainty. *J. Climate*, **35**, 6827–6842, https://doi.org/10.1175/JCLI-D-21-0958.1.

Ye, K., and G. Messori, 2021: Inter-model spread in the wintertime Arctic amplification in the CMIP6 models and the important role of internal climate variability. *Global Planet. Change*, **204**, 103543, https://doi.org/10.1016/j.gloplacha.2021.103543.

Yeager, S. G., A. R. Karspeck, and G. Danabasoglu, 2015: Predicted slowdown in the rate of Atlantic sea ice loss. *Geophys. Res. Lett.*, **42**, 10 704–10 713, https://doi.org/10.1002/2015GL065364.

Zanowski, H., A. Jahn, and M. M. Holland, 2021: Arctic Ocean freshwater in CMIP6 ensembles: Declining sea ice, increasing ocean storage and export. *J. Geophys. Res. Oceans*, **126**, e2020JC016930, https://doi.org/10.1029/2020JC016930.

Zhang, R., 2015: Mechanisms for low-frequency variability of summer Arctic sea ice extent. *Proc. Natl. Acad. Sci. USA*, **112**, 4570–4575, https://doi.org/10.1073/pnas.1422296112.