Assessment and Ranking of Climate Models in Arctic Sea Ice Cover Simulation: From CMIP5 to CMIP6

ZILI SHEN,^{a,b} ANMIN DUAN,^{b,c} DONGLIANG LI,^a AND JINXIAO LI^b

^a *Key Laboratory of Meteorological Disaster, Ministry of Education/Joint International Research Laboratory of Climate and Environmental Change/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, China*

^b *LASG, Institute of Atmospheric Physics, Chinese Academy of Science, Beijing, China*

^c *CAS Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Science, Beijing, China*

(Manuscript received 23 April 2020, in final form 17 January 2021)

ABSTRACT: The capability of 36 models participating in phase 6 of the Coupled Model Intercomparison Project (CMIP6) and their 24 CMIP5 counterparts in simulating the mean state and variability of Arctic sea ice cover for the period 1979–2014 is evaluated. In addition, a sea ice cover performance score for each CMIP5 and CMIP6 model is provided that can be used to reduce the spread in sea ice projections through applying weighted averages based on the ability of models to reproduce the historical sea ice state. Results show that the seasonal cycle of the Arctic sea ice extent (SIE) in the multimodel ensemble (MME) mean of the CMIP6 simulations agrees well with observations, with a MME mean error of less than 15% in any given month relative to the observations. CMIP6 has a smaller intermodel spread in climatological SIE values during summer months than its CMIP5 counterpart. In terms of the monthly SIE trends, the CMIP6 MME mean shows a substantial reduction in the positive bias relative to the observations compared with that of CMIP5. The spread of September SIE trends is very large, not only across different models but also across different ensemble members of the same model, indicating a strong influence of internal variability on SIE evolution. Based on the assumptions that the simulations of CMIP6 models are from the same distribution and that models have no bias in response to external forcing, we can infer that internal variability contributes to approximately $22\% \pm 5\%$ of the September SIE trend over the period 1979–2014.

KEYWORDS: Arctic; Sea ice; Coupled models; Model comparison; Model evaluation/performance

1. Introduction

Sea ice plays a key role in the global climate system. It can change the surface albedo and prevent heat exchange between the ocean and atmosphere, acting as a barrier between them (Screen and Simmonds 2010; Screen and Blackport 2019; Serreze and Barry 2011; Stroeve et al. 2012a). As assessed from satellite data, Arctic sea ice has been declining in all months, and the largest declining trend occurs in September. The years 2012 and 2020 marked record lows for September Arctic sea ice extent (SIE), reaching values of $3.4 \times 10^6 \text{ km}^2$ and $3.92 \times 10^6 \text{ km}^2$, respectively (Parkinson and Comiso 2013). Fauria et al. (2010) and Kinnard et al. (2011) reconstructed pre-satellite records based on terrestrial proxies and found that the decline in Arctic sea ice over the last few decades has been unprecedented during the past 1450 years. Recent studies estimated that greenhouse gas forcing has contributed up to 50% of the observed September Arctic sea ice extent decline trend during the past three decades (Ding et al. 2017, 2019; Kay et al.

2011; Stroeve et al. 2007, 2012b; Zhang 2010). Moreover, the Arctic perennial (ice that survives the summer) and multiyear (ice that has survived at least two summers) SIEs decreased at rates of $11.5\% \pm 2.1\% \text{ decade}^{-1}$ and $13.5\% \pm 2.5\% \text{ decade}^{-1}$, respectively, during 1979–2012 (Comiso 2012; Stocker et al. 2013; Parkinson and Comiso 2013), resulting in a transition toward a younger and thinner ice cover (Maslanik et al. 2007). The dramatic loss of sea ice has given rise to the question of when a seasonally ice-free Arctic Ocean will occur, the estimation of which has been highlighted as a “grand challenge” in the scientific community.

Climate models are important tools for providing insight into the key processes governing the observed sea ice loss and to estimate when a seasonally ice-free Arctic might be witnessed. Evaluation of climate model simulations against observations allows for better confidence in their future sea ice projections and understanding of the potential shortcomings of their performance. The Coupled Model Intercomparison Project (CMIP) archive is composed of a large quantity of state-of-the-art coupled atmosphere–ocean and Earth system models and provides a platform for assessing the capability of different climate models to simulate the Arctic sea ice cover from various perspectives (e.g., mean state, trends, interannual variability). The sea ice data from phase 3 of CMIP (CMIP3), which were used in the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC), has been assessed by many researchers (e.g., Rampal et al. 2011; Stroeve et al. 2007; Winton 2011; Zhang 2010). Based on these studies, it was found that the CMIP3 multimodel

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-20-0294.s1>.

Corresponding author: Anmin Duan, amduan@lasg.iap.ac.cn

DOI: 10.1175/JCLI-D-20-0294.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

ensemble (MME) mean largely underestimated the decrease in September Arctic SIE over the past several decades, and almost none of the individual CMIP3 model simulations could produce declining trends as strong as in the observations. In CMIP5, the models showed a large improvement, with a quarter of all individual model ensemble members showing a September SIE trend as strong as, or stronger, than in the observations over the satellite era (Stroeve et al. 2012a). Some studies have pointed out that new sea ice albedo parameterizations in CMIP5 allowing for melt ponds, as well as several models characterized by a rather thin winter sea ice cover, were responsible for the improvement in the SIE trend from CMIP3 to CMIP5 (Holland et al. 2012; Pedersen et al. 2009; Stroeve et al. 2012a). Nonetheless, underestimation of the sea ice decline, which was a feature of CMIP3 models, remains in CMIP5.

The different sensitivities of sea ice to global warming between models and observations might be an important factor in the discrepancy between observed and simulated sea ice trends (Mahlstein and Knutti 2012; Notz and Stroeve 2016; Rosenblum and Eisenman 2017). Moreover, internal variability may act as another possible explanation for these differences (Deser et al. 2016; Ding et al. 2017; Kay et al. 2011; Overland and Wang 2007; Screen and Francis 2016; Screen 2018). Ding et al. (2017) suggested that the internal variability of Arctic summer circulation trend contributes around 50% to the September sea ice loss and this atmospheric circulation over the Arctic has a strong link to the tropical SST variability (Ding et al. 2014). The failure of climate models to correctly capture the observed Arctic and tropical linkage may be responsible to a large extent for the discrepancy between modeled and observed September sea ice decline on multidecadal time scales (Ding et al. 2019). Other researchers have also proposed a link between Atlantic heat transport and regional internal variability of the winter Barents Sea SIE (Årthun et al. 2012; Li et al. 2017; Smedsrud et al. 2013). Problems with simulating high-latitude winds, ocean heat advection and mixing, and poleward atmospheric heat transport, which could lead to a nonlinear sea ice decrease, might also be causes of the mismatch between modeled and observed results (Koldunov et al. 2010; Melsom et al. 2009; Notz et al. 2013). The variety of possible biases does not allow us to rank the performances of models based on one single parameter. This is why, in this study, we select several metrics to give a synthesis of the sea ice cover performance of each model.

In this study, we evaluate the performance of climate models contributing to CMIP6 in simulating the Arctic sea ice cover. Unlike Notz et al. (2020), which mainly focused on the mean state and future evolution of the Arctic sea ice area (SIA) and volume in March and September, we focus on the ability of models to simulate both the SIE and SIA from different aspects (e.g., seasonal cycle, internal variability, trend) to provide a more comprehensive picture of sea ice cover performance. We compare the results of CMIP6 models to their CMIP5 counterparts, as well as against observations, and give each model a quantitative score based on their sea ice cover representation, thus enabling each modeling center to have a general picture of their strengths and weaknesses in sea ice

cover simulation and to identify any improvement compared with its previous model version. The large intermodel spread in the projected evolution of sea ice poses a significant challenge for stakeholders and policy makers in their adaptive planning (Liu et al. 2013; Overland and Wang 2007; Stroeve and Notz 2015). The credibility of a climate model will be increased if the model can well reproduce the past observed climate. Therefore, the score developed here can also be used as an approach to reducing the spread in sea ice projection through assigning weighted averages based on the ability of models to reproduce the historical sea ice state (Mahlstein and Knutti 2012; Overland and Wang 2013; Stroeve et al. 2007, 2012b; Wang and Overland 2012). On the other hand, the timing of when a September ice-free Arctic occurs is critical for marine activities and ecosystems, and hence is a major focus of both scientists and stakeholders. However, internal variability could potentially overwhelm the signal of anthropogenic forcing on decadal time scales or so; and in this respect, estimation of the relative contributions of internal variability and anthropogenic forcing to the observed summer sea ice loss is given in this study, which is of great importance for decision-makers when interpreting the robustness of sea ice projections.

We present the data sources and methods for calculating SIE in section 2. In section 3, we assess the climatological, seasonal cycle, internal variability, and long-term trends of SIE. In section 4, we show the sea ice concentration (SIC) distribution for a closer inspection of the geophysical distribution of sea ice cover. Section 5 presents a synthesis of SIE and SIA representations in each model. The observed and modeled September SIE trend uncertainty and an estimation of the contribution of internal variability to the September SIE trend are presented in section 6. Finally, a discussion and conclusions are given in section 7.

2. Datasets and methods

a. Observations

The sea ice observational data are from the U.S. National Snow and Ice Data Center (NSIDC; <http://nsidc.org/data/seaice/>). To provide a complete picture of the range of available observational products and their associated uncertainties, we use a number of different observational records of SIC obtained with different algorithms: 1) NASA Team (Cavalieri et al. 1996), 2) Bootstrap (Comiso 2017), and 3) a synthesis of the Bootstrap and NASA Team [the NOAA/NSIDC climate data record (CDR) of passive microwave sea ice concentration; Meier et al. 2017]. When comparing the modeled spatial distribution of sea ice cover to the observed sea ice concentration, only CDR is used. For monthly values of SIE and SIA, we use data from three different datasets: 1) Bootstrap, 2) NASA Team, and 3) the NSIDC sea ice index (version 3) (Fetterer et al. 2017).

b. Model outputs

Model outputs are based on the Global Climate and Earth System Model simulations from the CMIP6 data archive on the Earth System Grid Federation (ESGF) website (<https://esgf-node.llnl.gov/search/cmip6/>). We use 181 simulations for the

period of 1979–2014 from 36 CMIP6 models, employing the historical experiments (1850–2014) (Table 1). Most models offer at least three ensemble members as part of their historical simulation, while some provide only one. The only difference among different ensemble members from the same model is the randomly chosen initial conditions provided for the year 1850.

When comparing the simulated Arctic SIE and SIA with observations, we use the period of 1979–2014 due to the availability of satellite observations. To better understand the improvements and shortcomings of the CMIP6 models compared with those of CMIP5, we select 24 CMIP5 models (Table 1) for comparison. We augmented the CMIP5 twentieth-century historical simulations with results from the CMIP5 representative concentration pathway 4.5 (RCP4.5) future emissions scenario simulations to cover the period 1979–2014, in order to achieve the same time length as that of the CMIP6 historical simulations. RCP4.5 stabilizes radiative forcing at 4.5 W m^{-2} at the end of the twenty-first century (Meehl et al. 2012). The RCP4.5 pathway was selected from the four available future emissions scenarios to represent the period after 2005, noting that the choice of RCP scenario has minimal influence during 2006–14 (Li et al. 2017). The 24 selected CMIP5 models were either the previous generation of each CMIP6 model used in this study or were produced by the same modeling center.

c. Methods

The SIE and SIA were computed using the native grid of each model. The nearest neighbor interpolation method was used to interpolate SIC outputs from models with different horizontal resolutions to the satellite grid for comparison of SIC between observations and models. The MME mean was calculated with equal weight for each climate model. The result of each model is represented by the mean of all the ensemble members provided by this individual model in order to eliminate random errors. To calculate the ensemble mean trend and variability of the metrics of SIA and SIE for each model, we used the following procedure: trend and internal variability are calculated for each member individually and then averaged over all ensemble members of that model. The same procedure is followed to compute the MME mean trend of each grid cell sea ice concentration. For models with more than 10 ensemble members, we only chose the first 10 members to avoid internal variability being eliminated by the mean of large ensemble members. Since some models only provide one ensemble member, we cannot calculate the internal variability from the ensemble spread of these models. As an alternative, following the technique used by Thompson et al. (2015) and England et al. (2019), the internal variability in observations and models is calculated as the standard deviation (SD) of the detrended SIE over the period of 1979–2014 based on the assumption that the variations in SIE on decadal time scales account for a relatively small part of the total variance in SIE. One intermodel SD between all the models is used as the metric to quantify the uncertainty in model simulations. The model SIE is calculated by summing the area of all grid cells with an SIA fraction exceeding 15%, which is the widely recognized threshold for dividing the sea ice boundary in passive microwave SIC products (Parkinson et al. 1999).

In contrast with the Notz et al. (2020) sea ice evaluation study, we selected SIE over SIA as the primary measure of sea ice cover. The difference in the climatological SIE between the two algorithms (NASA Team and Bootstrap) is $0.27 \times 10^6 \text{ km}^2$ in March and $0.49 \times 10^6 \text{ km}^2$ in September. For SIA, the difference is $1.06 \times 10^6 \text{ km}^2$ in March and $1.21 \times 10^6 \text{ km}^2$ in September. The larger difference in SIA than in SIE between the two algorithms (Fig. S1 in the online supplemental material; Notz et al. 2013) is the decisive reason why we chose SIE as the primary measure of sea ice cover in this research. Because there are also some shortcomings and limitations in SIE owing to its grid dependence and nonlinearity, which could mislead a model evaluation based solely on SIE (Notz 2014), we also consider the SIA performance in section 5 when synthesizing the sea ice cover performance for each model.

3. Monthly Northern Hemisphere sea ice cover

a. Climatology

Following the method used in Stroeve et al. (2012a), we first compare the seasonal cycle of the observed SIE with that simulated by climate models (Fig. 1). The seasonal transition between sea ice melting and freezing is crucial for heat exchange and kinetic energy transfer (Arzel et al. 2006; Döscher et al. 2014; Francis et al. 2009; Stroeve et al. 2007, 2012a). A well-simulated seasonal cycle is, to some extent, related to the realistic sensitivity to solar forcing in a model.

Both CMIP5 and CMIP6 can effectively reproduce the seasonal cycle of Arctic SIE with an MME mean error less than about 15% in every month (Fig. 1). The timing of the maximum and minimum SIE in the models is similar to those in the satellite records. The mean of the MME seasonal cycle of SIE from the 36 CMIP6 models has a positive bias relative to the observations in all months (Fig. 1). The magnitude of this positive deviation increases from $0.44 \times 10^6 \text{ km}^2$ in February to a peak of around $1.12 \times 10^6 \text{ km}^2$ in May, and then decreases sharply until July. From August to December, the positive deviation fluctuates periodically around $0.24 \times 10^6 \text{ km}^2$. July, August, October, and November are the four months that have their MME means falling within the one standard deviation of the three satellite products. Compared with CMIP5, the CMIP6 SIE shows the most visible improvement in July and August, whereas for other months, the bias is not reduced or is even larger. Moreover, the same conclusions can be drawn when using SIA, except for the fact that the SIA bias is larger for the CMIP6 MME mean than for the CMIP5 MME mean in all months (Fig. S2).

The intermodel spread of SIE simulated by CMIP5 and CMIP6 is relatively larger in winter compared with other months. Compared with CMIP5, the interquartile ranges are lower in CMIP6, especially in August and September.

In addition to the general MME features of SIE, we present SIE results from CMIP5 to CMIP6 for each modeling center. In March, both CMIP5 and CMIP6 have a symmetric distribution of MME spread around the observed values for SIE (Fig. 2a). For most modeling centers, the bias in the climatological March SIE relative to observations is reduced in CMIP6 compared with that in CMIP5, but three modeling centers [the

TABLE 1. Details of the CMIP5 and CMIP6 models, grouped by modeling center, used in this study.

Modeling center number	Modeling center	Nation	CMIP5		CMIP6	
			Model name	No. of ensemble members	Model name	No. of ensemble members
1	Australian Research Council Centre of Excellence for Climate System Science (CSIRO-ARCCSS)	Australia	ACCESS1.0	2	ACCESS-CM2 (https://doi.org/10.22033/ESGF/CMIP6.4271)	3
			ACCESS1.3	2	ACCESS-ESM1.5 (https://doi.org/10.22033/ESGF/CMIP6.4272)	9
2	Beijing Climate Center, China Meteorological Administration (BCC)	China	BCC_CSM1.1	3	BCC_CSM2-MR (https://doi.org/10.22033/ESGF/CMIP6.2948)	2
					BCC-ESM1 (https://doi.org/10.22033/ESGF/CMIP6.2949)	3
3	Chinese Academy of Meteorology Science (CAMS)	China	—	—	CAMS-CSM1.0 (https://doi.org/10.22033/ESGF/CMIP6.9754)	1
4	Canadian Centre for Climate Modeling and Analysis (CCCma)	Canada	CanCM4	10	CanESM5 (https://doi.org/10.22033/ESGF/CMIP6.10260)	10
5	Community Earth System Model Contributors (NCAR)	United States	CanESM2	5	CESM2 (https://doi.org/10.22033/ESGF/CMIP6.7627)	10
			CESM1-CAM5	3		
			CESM1-WACCM	4	CESM2-WACCM (https://doi.org/10.22033/ESGF/CMIP6.10071)	3
6	Centre National de Recherches Météorologiques (CNRM-CERFACS)	France	CNRM-CM5	10	CNRM-CM6.1 (https://doi.org/10.22033/ESGF/CMIP6.4066)	9
					CNRM-CM6.1-HR (https://doi.org/10.22033/ESGF/CMIP6.4067)	1
					CNRM-ESM2.1 (https://doi.org/10.22033/ESGF/CMIP6.4068)	5
7	Lawrence Livermore National Laboratory (LLNL)	United States	—	—	E3SM-1.0 (https://doi.org/10.22033/ESGF/CMIP6.4497)	1
8	European Consortium (EC)	Europe	EC-EARTH	1	EC-Earth3 (https://doi.org/10.22033/ESGF/CMIP6.4700)	8
					EC-Earth3-Veg (https://doi.org/10.22033/ESGF/CMIP6.4706)	6
9	Chinese Academy of Science (CAS)	China	FGOALS-g2	1	FGOALS-f3-L (https://doi.org/10.22033/ESGF/CMIP6.3355)	3
					FGOALS-g3 (https://doi.org/10.22033/ESGF/CMIP6.3356)	6
10	NOAA/Geophysical Fluid Dynamics Laboratory (GFDL)	United States	GFDL-CM3	5	GFDL-CM4 (https://doi.org/10.22033/ESGF/CMIP6.8594)	1
			GFDL-ESM2G	1	GFDL-ESM4 (https://doi.org/10.22033/ESGF/CMIP6.8597)	1
			GFDL-ESM2M	1		

TABLE 1. (Continued)

Modeling center number	Modeling center	Nation	CMIP5		CMIP6	
			Model name	No. of ensemble members	Model name	No. of ensemble members
11	NASA Goddard Institute for Space Studies (GISS)	United States	GISS-E2-R	10	GISS-E2.1-H (https://doi.org/10.22033/ESGF/CMIP6.7128)	10
			GISS-E2-H	8	GISS-E2.1-G (https://doi.org/10.22033/ESGF/CMIP6.7127)	10
12	Met Office Hadley Centre (MOHC)	United Kingdom	HadGEM2-ES	5	HadGEM3-GC31-LL (https://doi.org/10.22033/ESGF/CMIP6.6109)	4
			HadGEM2-CC	3	HadGEM3-GC31-MM (https://doi.org/10.22033/ESGF/CMIP6.6112)	2
			—	—	UKESM1.0-LL (https://doi.org/10.22033/ESGF/CMIP6.6113)	4
13	L'Institut Pierre-Simon Laplace (IPSL)	France	IPSL-CM5A-LR	6	IPSL-CM6A-LR (https://doi.org/10.22033/ESGF/CMIP6.5195)	10
			IPSL-CM5B-LR	1		
14	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (University of Tokyo), and National Institute for Environmental Studies (MIROC)	Japan	MIROC5	5	MIROC6 (https://doi.org/10.22033/ESGF/CMIP6.5603)	10
			MIROC-ESM-CHEM	1	MIROC-ES2L (https://doi.org/10.22033/ESGF/CMIP6.5602)	3
15	Max Planck Institute for Meteorology (MPI)	Germany	MPI-ESM-MR	3	MPI-ESM-1.2-HAM (https://doi.org/10.22033/ESGF/CMIP6.5016)	2
			—	—	MPI-ESM1.2-HR (https://doi.org/10.22033/ESGF/CMIP6.6594)	7
			—	—	MPI-ESM1.2-LR (https://doi.org/10.22033/ESGF/CMIP6.6595)	10
16	Meteorological Research Institute (MRI)	Japan	MRI-CGCM3	5	MRI-ESM2.0 (https://doi.org/10.22033/ESGF/CMIP6.6842)	5
17	Nanjing University of Information Science and Technology (NUIST)	China	—	—	NESM3 (https://doi.org/10.22033/ESGF/CMIP6.8769)	4
18	Norwegian Climate Centre (NCC)	Norway	NorESM1-ME	1	NorCPM1 (https://doi.org/10.22033/ESGF/CMIP6.10894)	10
			—	—	NorESM2-LM (https://doi.org/10.22033/ESGF/CMIP6.8036)	3
			—	—	NorESM2-MM (https://doi.org/10.22033/ESGF/CMIP6.8040)	1
19	Seoul National University (SNU)	South Korea	—	—	SAM0-UNICON (https://doi.org/10.22033/ESGF/CMIP6.7789)	1
20	Research Center for Environmental Changes, Academia Sinica (AS-RECE)	China	—	—	TaiESM1 (https://doi.org/10.22033/ESGF/CMIP6.9755)	1

FIG. 1. Comparison of the CMIP5 (blue) and CMIP6 (red) MME mean seasonal cycle with the observed SIE seasonal cycle for the period 1979–2014. The CMIP5 and CMIP6 MMEs are shown as box-and-whisker plots. For models with more than one ensemble member, the mean of the ensemble members is calculated. Boxes represent interquartile ranges (25th–75th percentiles). Median (50th percentile) and mean extents are shown by the thick horizontal bar and dot in each box, respectively. Whiskers represent the 10th and 90th percentiles. Extents less than the 10th percentile or greater than the 90th percentile are shown as individual crosses. The thick solid line denotes the average of three available satellite products (NASA Team, Bootstrap, and NSIDC-CDR). The gray shading denotes one standard deviation of the three satellite products. Note that this figure is comparable to Fig. 1 in [Stroeve et al. \(2012a\)](#), but here we use CMIP6 data.

Chinese Academy of Meteorology Science (CAMS), Lawrence Livermore National Laboratory (LLNL), and GISS] largely overestimate the March SIE in CMIP6 (bias relative to observations larger than 30%) ([Fig. 2a](#)).

In September, the situation is different. The MME spread is less evenly distributed around the observations in CMIP6 compared to that in March ([Fig. 2b](#)). Some CMIP6 models (FGOALS-g3, GISS-E2-H and NorCMP1) have a bias of more than 50% relative to observations, resulting in a larger bias in the CMIP6 MME mean of the September-mean SIE compared with that of CMIP5 (the bias is 7.2% and 0.46% for CMIP6 and CMIP5, respectively; [Fig. 2b](#)).

For the seasonal cycle amplitude of SIE, which is defined as the difference between the March and September SIEs, there is a slight improvement in the CMIP6 MME mean compared to that of CMIP5; the average modeled SIE amplitude in CMIP6 ($9.14 \times 10^6 \text{ km}^2$) is comparable to the observations ($9.09 \times 10^6 \text{ km}^2$) ([Fig. 2c](#)). This metric can to some extent reflect the ability of models to capture the seasonal transition of SIE. On the other hand, it should be noted that a good match in the seasonal cycle amplitude does not mean the model has no bias in both its March and September SIEs; it may result from the error compensation in the September and March SIEs. The

presence of a large bias in the simulated seasonal cycle might indicate some shortcomings in models' sensitivity to solar forcing or to increasing greenhouse gas associated with long-wave radiative flux (e.g., water vapor and cloud feedbacks). It is worth noting that among the CMIP6 models, six of them (ACCESS-ESM1.5, BCC-ESM1, CNRM-CM6.1-HR, CanESM5, MPI-ESM1.2-HR, and MRI-ESM2.0) show a seasonal cycle amplitude very close to the observations.

b. Internal variability and evolution

In this section, several metrics are selected to assess the variability of the simulated sea ice in the CMIP5 and CMIP6 models. The modeled internal variability is largely underestimated in the CMIP6 models based on the annual mean, March and September values compared with the observations ([Figs. 3a–c](#)). However, it is interesting to find that in CMIP5 models, the opposite is the case (except September). For the annual mean and March values, the distribution of internal variability in SIE for CMIP5 and CMIP6 is in both cases asymmetric around the observations, with around 86% of CMIP6 models underestimating the internal variability in SIE and 92% of CMIP5 models overestimating this value ([Figs. 3a,b](#)). In September, the CMIP5 MME mean of internal

FIG. 2. SIE climatology metrics for historical simulations in CMIP5 (blue) and CMIP6 (red): 1979–2014 mean SIE in (a) March and (b) September (10^6 km^2); (c) seasonal cycle amplitude of SIE, defined as the difference between the March and September SIE. The CMIP5 and CMIP6 MMEs are shown as box-and-whisker plots on the left of each panel, with the boxes and whiskers representing interquartile ranges (25th–75th percentiles) and one intermodel standard deviation, respectively. Median (50th percentile) and mean values are shown by the thick horizontal bar and dot in each box, respectively. Individual CMIP5 models are shown as filled gray circles, and individual CMIP6 models are listed in the legend. The black dashed line denotes the average of available satellite products (NASA Team, Bootstrap, and NSIDC-CDR). The gray shading denotes one standard deviation of the three satellite products.

variability is comparable to the observations (Fig. 3c). However, for the CMIP6 MME mean, the underestimation in September is even larger than that in March and the annual mean.

In terms of RMSE, there is little difference between the CMIP5 and CMIP6 MME mean in the annual mean and for

March (Figs. 3d,e). In particular, some models show annual mean (e.g., CNRM-ESM2.1 and TaiESM1) and March (e.g., ACCESS-ESM1.5 and IPSL-CM6A-LR) SIE values that are very close to the observed one. For September, the CMIP6 MME mean exhibits an improvement compared to CMIP5. However, given the short time series and insufficient number of ensemble members, we may not expect any individual model with limited ensemble members to have their modeled time series match the observed one perfectly. Thus, the results presented above are for reference only and the ability of models to reproduce these metrics must be assessed with caution.

Overall, it is not possible for one model to stand out above all others in all metrics, as each has strengths and weaknesses. However, the results of some metrics are clearly linked—for example, the monthly SIE climatological metrics are significantly correlated with the SIE RMSE in that month.

c. Trend

Trends in modeled SIE are often used as a measure of a model's ability to capture the response of the sea ice to the imposed changes in external forcing (Stroeve et al. 2012a). The linear trends are obtained according to the standard least squares linear method. Figure 4 shows the monthly trends of observed SIE along with the CMIP5 and CMIP6 MME means for the period 1979–2014. The observed Arctic SIE decreases in every month of the year, and the declining trend in each month is significant at the 95% level. The largest observed SIE decline occurs in the period from August to October, with the greatest decline in September ($-0.87 \times 10^6 \text{ km}^2 \text{ decade}^{-1}$).

Box-and-whisker plots show that both the CMIP5 and CMIP6 models have a very large spread of 36-yr trends in August and September, with September having the largest intermodel range (SD of $0.25 \times 10^6 \text{ km}^2 \text{ decade}^{-1}$ for CMIP5 and $0.33 \times 10^6 \text{ km}^2 \text{ decade}^{-1}$ for CMIP6). The intermodel differences in the monthly trends are relatively small in the winter months in CMIP6, with March having the smallest spread (SD of $0.15 \times 10^6 \text{ km}^2 \text{ decade}^{-1}$). In CMIP5, the smallest intermodel spread occurs in June (SD of $0.12 \times 10^6 \text{ km}^2 \text{ decade}^{-1}$).

The absolute value of the mean and median SIE trends of the CMIP6 ensemble are underestimated in most months with respect to the observations. Although the distributions of the CMIP5 SIE trends are similar to those of CMIP6, the mean and median trends of CMIP5 have a positive bias relative to both the CMIP6 model results and the observations.

In March, CMIP5 underestimates the SIE decline trend by 37%, versus 11% for CMIP6. An inspection of the individual modeling centers reveals that only 36% of the CMIP6 models, versus 17% of the CMIP5 models, fall within a 15% error of the observed value ($-0.39 \times 10^6 \text{ km}^2 \text{ decade}^{-1}$) of the SIE trend in March (Fig. 5a). In particular, some CMIP6 models (e.g., EC-Earth3-Veg, GFDL-ESM4, GISS-E2.1-H, MPI-ESM1.2-HR, and MPI-ESM1.2-LR; see Fig. 5a) show a March trend that is very close to the observed one.

In September, the large intermodel spread in the trends of the CMIP6 models (SD of $0.32 \times 10^6 \text{ km}^2 \text{ decade}^{-1}$) is mainly caused by the large bias in CAMS-CSM1.0, E3SM1.0, FGOALS-g3, NESM3, and NorCMP1, with the SIE trend bias

FIG. 3. SIE variability metrics for historical simulations in CMIP5 (blue) and CMIP6 (red): standard deviation of detrended 1979–2014 (a) annual, (b) March, and (c) September SIE (10^6 km^2); (d) annual, (e) March, and (f) September SIE RMSE (10^6 km^2). The CMIP5 and CMIP6 MMEs are shown as box-and-whisker plots on the left of each panel, with the boxes and whiskers representing interquartile ranges (25th–75th percentiles) and one inter-model standard deviation, respectively. Median (50th percentile) and mean values are shown by the thick horizontal bar and dot in each box, respectively. Individual CMIP5 models are shown as filled gray circles, and individual CMIP6 models are listed in the legend. The black dashed line denotes the average of available satellite products (NASA Team, Bootstrap, and NSIDC-CDR). The gray shading denotes one standard deviation of the three satellite products.

being more than 70% relative to observations (Fig. 5b). The underestimation of the SIE trend for the CMIP5 MME mean is 32%, and for CMIP6, it is 22%. Additionally, 22% of the CMIP6 models versus 12% of the CMIP5 models have a September SIE trend error within 15% of the observations.

In summary, CMIP6 models are on average closer to observations than CMIP5 in terms of the SIE trend. This may be partly related to the more realistic simulation of sea ice sensitivity (defined as the ratio of sea ice loss to a given amount of global warming) in CMIP6 as indicated by Notz et al. (2020). However, this change in sea ice sensitivity or SIE trend could

also be related to the higher value in historical ozone radiative forcing and black carbon emissions in CMIP6 relative to CMIP5 rather than the improvement in model physics (Checa-Garcia et al. 2018; Gidden et al. 2019).

4. Spatial distribution of the sea ice concentration

Cavalieri and Parkinson (2012) pointed out that the Arctic sea ice trends over the period 1979–2010 display spatial inhomogeneity, which is indicative of the complex nature of the Arctic climate system. Thus, we also need to pay attention to

FIG. 4. As in Fig. 1, but for monthly trends of SIE ($10^6 \text{ km}^2 \text{ decade}^{-1}$).

the spatial pattern of SIC to avoid overconfidence in model performance and to exclude the compensation of errors of opposite signs in different regions when summing up areas for all grid cells with SIC greater than 15% to calculate SIE. In this section, we only discuss the MME mean features of SIC (the performances of individual models can be inspected in Figs. S3 and S4).

Both the CMIP5 and CMIP6 MME means for the March SIE display reasonable agreement with observations (Fig. 1). However, when we look at the spatial pattern as shown in Figs. 6a and 6c, we find this not to be the case regionally. The CMIP5 and CMIP6 MME means overestimate the SIC values by approximately 35% in the Barents Sea Opening (BSO) and Greenland Sea. For September, the underestimation of the SIC in the Arctic central region relative to the observed SIC in the CMIP5 MME mean is slightly narrowed in CMIP6 MME mean, which might be related to the improved melt pond schemes applied in the CMIP6 models (Figs. 6b,d).

In terms of the SIC trend, the observed March SIC trend displays spatial inhomogeneity, with the most pronounced sea ice declining trend in the northeast section of the Barents Sea (Fig. 7a). In contrast to the observations, the modeled March Arctic SIC response to CMIP5 and CMIP6 external forcing displays a much weaker decrease (trend underestimation of approximately 20%) in the regions where the observed decline is most pronounced (Figs. 7c,e). Li et al. (2017) found that Atlantic heat transport through the BSO (HT_{BSO}) in CMIP5 forced simulations is too small to have caused the observed Barents Sea ice decline. Based on CMIP5 unforced control simulations, they found that current climate models have an unrealistically small amplitude of low-frequency variability in HT_{BSO} . Additionally, England et al. (2019) found that internal variability is responsible for more than 75% of the recent observed March sea ice changes. Therefore, the failure of models to capture the realistic amplitudes of low-frequency internal variability in HT_{BSO} possibly causes the discrepancy between observed and simulated winter SIC declining trend in the BSO region in March.

The observed September SIC trends are dominated by the decrease in sea ice in marginal ice zones from the Kara Sea

eastward to the Beaufort Sea, especially in the Beaufort, Chukchi, and East Siberian Seas, featuring an average decline of more than $25\% \text{ decade}^{-1}$ (Fig. 7b). We can see that both the CMIP6 and CMIP5 models have a general tendency to underestimate the decrease in SIC alongside the marginal ice zones mentioned above, especially in the Arctic Pacific Sector (APS; Figs. 7d,f). The CMIP5 and CMIP6 MME means exhibit a similar pattern for the September SIC trend but differ in magnitude with the CMIP6 MME mean exhibiting a slightly (approximately 8%) stronger rate of decline in the Chukchi and East Siberian Seas compared with that of CMIP5. England et al. (2019) estimated the internal variability accounted for around 35% of the recent observed September sea ice decline. Unlike the observed March sea ice change, of which internal variability account for more than 75%, failure of the models to capture the observed magnitude of the September sea ice decline is not primarily caused by internal variability, rather, it is possibly the model physics, such as the cloud cover representation or the treatment of melt ponds (Holland et al. 2012; Roeckner et al. 2012). However, the larger observational uncertainty in summer due to the difficulties for satellites to get the sea ice cover correct when the sea ice is covered with water from melt ponds may also cause the discrepancy between the observed and modeled September SIC.

5. Sea ice cover error scores

This section gives a synthesis of the sea ice cover simulation performance for all CMIP5 and CMIP6 models. To quantify how well the models reproduce the basic characteristics of sea ice cover, we use the subset of six metrics presented in the previous sections: the climatological state, linear trend and internal variability of March and September SIE. Despite the smaller uncertainty in SIE than in SIA between two algorithms, and many studies having used SIE as the sole variable to evaluate the quality of modeled sea ice cover (Li et al. 2017; Liu et al. 2013; Stroeve et al. 2007, 2012a), SIE still has its limitations owing to its non-linearity and cancelling biases in SIC, as suggested by Notz (2014). As discussed in section 4, a good match between models and the observations on SIE does not mean that they have a good agreement on SIC. This is because when we calculate the SIE, all the grid cells with a SIC greater than 15% are counted as being sea ice covered. Hence, the differences between the observations and models at SIC larger than this threshold will vanish for the calculation of SIE. Therefore, notwithstanding the larger observational uncertainty, we also add SIA—a more direct integrative measure—to give a more precise description of SIC. Thus, the choice of which variable to represent the ability of the models to correctly simulate the sea ice cover can depend on one's specific interest. Following the method used by Bellenger et al. (2014), for each metric m and each model i , e_i is derived from the absolute difference between the modeled (m_i) and the observed (m_{ref}) metrics, divided by the CMIP5 + CMIP6 intermodel standard deviation ($\sigma_{\text{CMIP5+CMIP6}}$) as follows:

$$e_i = \frac{|m_i - m_{\text{ref}}|}{\sigma_{\text{CMIP5+CMIP6}}}. \quad (1)$$

FIG. 5. SIE trend metrics for historical simulations in CMIP5 (blue) and CMIP6 (red): 1979–2014 SIE trends in (a) March and (b) September ($10^6 \text{ km}^2 \text{ decade}^{-1}$). The CMIP5 and CMIP6 MMEs are shown as box-and-whisker plots on the left of each panel, with the boxes and whiskers representing interquartile ranges (25th–75th percentiles) and one intermodel standard deviation, respectively. Median (50th percentile) and mean values are shown by the thick horizontal bar and dot in each box, respectively. Individual CMIP5 models are shown as filled gray circles, and individual CMIP6 models are listed in the legend. The black dashed line denotes the average of available satellite products (NASA Team, Bootstrap, and NSIDC-CDR). The gray shading denotes one standard deviation of the three satellite products.

The final SIE and SIA error score for each model is defined as the average of these selected metrics' scores; and the lower the SIE and SIA error score, the better the model performs in terms of the basic features of SIE and SIA. The separate

model ranking according to SIE and SIA score can refer to Fig. S5 in detail.

The average SIE score for the CMIP6 models is 0.88 (SD of 0.39) (Fig. 8b). This value is similar to the average score of

FIG. 6. The climatological (left) March and (right) September sea ice concentration during 1979–2014 with departure from NSIDC-CDR (model minus observations) for the (a),(b) CMIP5 and (c),(d) CMIP6 MME mean.

0.87 for CMIP5 (SD of 0.40). For the SIA score, the bias relative to the observations is larger than that for SIE in the CMIP6 models (Fig. 8b), with an average SIA score of 1.16 (SD of 0.54). The top four models (10th percentile) among the 36 CMIP6 models in terms of SIE scores are GFDL-CM4, TaiESM1, HadGEM3-GC31-LL, and EC-Earth3-Veg. In terms of SIA scores, the top four models are MPI-ESM-1.2-HAM, CNRM-CM6.1-HR, NorESM2-MM, and EC-Earth3-Veg. Among all the CMIP5 and CMIP6 models, EC-Earth3-Veg is the outstanding performer regarding sea ice cover. The overall larger bias relative to observations for SIA compared with that for SIE is mainly caused by the poor representation in the value of the SIC in models, as shown in Figs. 6 and 7.

Since the credibility in the models' projections of future climate is established on their demonstrated ability to reproduce the past

climate, the combination of SIE and SIA can provide a thorough estimate of a model's performance in sea ice cover simulation and score-based ranking can be used as an approach to reduce the spread and correct the bias in the projected time of an ice-free Arctic based on the model's ability to reproduce the historical sea ice cover state, which means giving higher weights to those models that provide more realistic simulations of historical sea ice when calculating the MME mean. Here, we provide two simple practical examples of how to use this method, as follows:

a. Model selection

One of the simplest but most transparent choices regarding “weighting” methods is model selection (inclusion or exclusion), that is, giving “outlier” models weights of 0 and selected models weights of 1 in future projection. In this application, we eliminate

FIG. 7. Sea ice concentration trends in (left) March and (right) September for the period 1979–2014 for (a),(b) NSIDC-CDR, (c),(d) CMIP5, and (e),(f) CMIP6 runs. Blue boxes indicate the APS (70° – 85° N, 110° E– 110° W) on the right and BSO (70° – 81° N, 15° – 60° E) on the left.

FIG. 8. Performance of SIE and SIA simulations for each model of (a) CMIP5 and (b) CMIP6. Six primary scores are used to depict the SIE quality: the normalized error (no units; a zero value indicates a perfect agreement with observations) of the climatological state, linear trend, and internal variability of March and September SIE and SIA between models and NSIDC satellite data. A diagonal line divides each cell into two parts, with the upper triangle representing March scores and the lower triangle representing September scores. In the bottom line (the SCORE line), the SIE score (upper triangle) is defined as the average of these March and September SIE scores, and the SIA score (lower triangle) is defined as the average of these March and September SIA scores.

the models that seriously fail to meet the observed mean and trend of September SIE (Wang and Overland 2012; Liu et al. 2013). Based on the SIE error score, we exclude models with September SIE score larger than 1.0, which means models with biases relative to the observations greater than the spread in the multimodel simulated changes are culled. This selection process results in the retention of 12 of 36 CMIP6 models. As shown in Fig. S6, the range of uncertainty in the projected September SIE by these 12 selected models has been largely reduced under both shared socioeconomic pathway (SSP) emission scenarios SSP2–4.5 and SSP5–8.5. The reduction is most obvious at the end of twenty-first century. The standard deviation among all models is 1.86×10^6 and $0.97 \times 10^6 \text{ km}^2$ for the period 2081–2100 under SSP2–4.5 and SSP5–8.5, respectively. This standard deviation is reduced to $0.61 \times 10^6 \text{ km}^2$ (SSP2–4.5) and $0.02 \times 10^6 \text{ km}^2$ (SSP5–8.5) when projections are composed of the 12 best-performing models, a reduction of 67% and 98% under SSP2–4.5 and SSP5–8.5, respectively. In addition, the models with reasonable mean and trend of September SIE relative to the observations project an earlier ice-free Arctic (defined as SIE in the Arctic less than 1 million km^2 for consecutive five years). Relative to the ensemble mean of all models, the reduced set of 12 models advance the timing of the ice-free Arctic by more than two decades under SSP5–8.5 (Fig. S6).

b. Nonuniform model weighting

In this method, each model is given a weight,

$$w_m = \frac{1}{e_m}, \quad (2)$$

where e_m is the sea ice cover error score in each model m . These weightings can be expressed as normalized quantities:

$$W_m = \frac{w_m}{\sum_m w_m}. \quad (3)$$

The weighted MME mean $S(t)$ can simply be expressed as follows:

$$S(t) = \sum_m W_m x_{m,t}, \quad (4)$$

where $x_{m,t}$ is the time series of sea ice cover for each model. This scheme thus minimizes the contribution of those models that perform poorly in the representation of Arctic sea ice. Here, we just provide a simple methodology; more complex weighting schemes could be applied if considering model independence (e.g., Giorgi and Mearns 2002; Sanderson et al. 2015, 2017). Using this model bias correction, we advance the timing for the September ice-free Arctic from 2070 to 2057 for the high-emission scenario (SSP5–8.5) (Fig. S7).

FIG. 9. September trends from 1979 to 2014 for all individual model ensemble members with confidence intervals (vertical lines). The 1σ and 2σ observed trends (using the NSIDC sea ice index) are shown in dark gray shading (1σ) and light gray shading (2σ). Note that this figure is comparable to Fig. 3 of [Stroeve et al. \(2012a\)](#), but here we use CMIP6 data.

It should be noted that although the scores used in this study can to a certain extent be recommended as an approach to rank the ability of models to simulate the sea ice cover, they should not be considered as precise and definitive. Since some models only provide one ensemble member, the large difference between the modeled and observed sea ice cover might be caused by the internal variability rather than the model error. Other studies have also shown that sea ice thickness and volume are important in controlling sea ice variability and uncertainties in projected sea ice trends ([Massonnet et al. 2018](#); [Mioduszewski et al. 2019](#)). Both the choice of metrics and the method used to define the model's scores can be revised depending on the specific research interest.

6. SIE trend uncertainty and estimation of the contribution of internal variability to the SIE trend

a. Level of agreement between the modeled and observed September SIE trend

Although the SIE trend bias relative to observations is narrowed in the CMIP6 MME mean compared with that of CMIP5, due to the existence of internal variability, it is critical to assess whether the simulated trends agree with the observations at a specified statistical significance level. As the observations are only one realization of how internal variability influences the SIE trend, it is not possible to expect an individual simulation to match well with the observational record on the decadal time scale, even for a perfect model,

but the observed value should fall within the range of results of the model's ensemble members ([Notz 2015](#); [Santer et al. 2008](#); [Stroeve et al. 2012a](#)). In this section, we show results for all ensemble members of the CMIP6 models, rather than calculating the ensemble mean for each model. Thus, we follow the lead of [Santer et al. \(2008\)](#) and evaluate whether the individual SIE trends of the CMIP6 model ensemble simulations are identical to the observed trend. The null hypothesis is that the modeled SIE trend of any given simulation of a participating model is consistent with the observed trend. The test statistic is

$$d = \frac{\beta_m - \beta_o}{\sqrt{\sigma_m^2 + \sigma_o^2}}, \quad (5)$$

where d is the standard error between the modeled and observed time series; β_m and β_o are the modeled and observed trends, respectively; and σ_m and σ_o represent the standard errors of the modeled and observed trend, respectively. In terms of the autocorrelations of the regression residuals, we calculate the effective sample size n_e :

$$n_e = n_1 \frac{1 - r_1}{1 + r_1}, \quad (6)$$

where n_1 is the total number of years taken to calculate the trend and r_1 is the lag-1 autocorrelation coefficient for the regression residuals. The autocorrelation coefficient for the observed 1979–

FIG. 10. Observed and CMIP5- and CMIP6-modeled linear trends of September Arctic SIE. Histograms show the distribution of CMIP5- and CMIP6-modeled trends, with the observed trend indicated by the green line. A Gaussian fit to each distribution is plotted in turquoise (CMIP5) and red (CMIP6). Cumulative percentages are shown as stepped lines.

2014 September SIE regression residual time series is 0.1, and for the modeled time series it ranges from 0.1 to 0.78.

Figure 9 shows the observed and modeled September trends from 1979 to 2014. The larger 2σ error bars of modeled trends indicate that the regression residual time series are more autocorrelated. Most ensemble members in CMIP6 underestimate the decreasing trend for September SIE, and the trends vary considerably from one ensemble member to another. This difference exists even for ensemble members from the same model. In total, 162 and 136 (including all 10 CNRM-CM6.1 ensemble members, all 5 FGOALS-g3 ensemble members, all 10 GISS-E2.1-H ensemble members, all 3 MIROC-ESM2L ensemble members, all 4 NESM3 ensemble members, and all 10 NorCPM1 ensemble members) out of 181 ensemble members have their simulated trends fall outside the 1σ and 2σ levels of the observed trend, respectively. Given the assumption that the trends in the observations and model ensemble members are from the same normal distribution with a mean of $\beta_m - \beta_o$ and an SD of $\sqrt{\sigma_m^2 + \sigma_o^2}$, then we can test whether the trend of the model time series is consistent with the observed trend. If the value of $|d|$ exceeds 1.96, then the null hypothesis is rejected at 95% confidence level. CNRM-CM6.1, FGOALS-g3, MIROC-ES2L, NESM3, and NorCPM1 are the models for which more than 75% of their ensemble members are significantly different from the observed trend at the 95% confidence level, which indicates that these models have nonnegligible problems in their model physics or sea ice sensitivities to changes in external forcing. Given the impact of internal variability on the trends and small ensemble size for some models (e.g., CAMS-CSM1.0), there could be trends simulated by models that differ markedly from the observed

trends, which is completely unrelated to model error. Thus, we cannot make a definitive judgement for those models with a small ensemble size without further analysis.

b. Contribution of internal variability

September is the month with the most rapid decline in sea ice, and the timing of an ice-free Arctic in September is a hot topic among both scientists and stakeholders. In some cases, internal variability can overwhelm the response of sea ice to changes in external forcing on interannual to decadal time scales, and a high contribution of internal variability might indicate a less predictable sea ice evolution in the future (DeRepentigny et al. 2020; Jahn 2018). Hence, understanding the contribution of internal variability to long-term trends is essential for decision-makers in their interpretations of the robustness of sea ice projections. To approximately quantify the relative role of internal variability versus anthropogenic forcing in the contribution to the September SIE decline, the distribution of the CMIP5 and CMIP6 simulated Arctic September SIE trend from 1979 to 2014 is shown in Fig. 10. The distribution reflects the possible influence of internal variability (e.g., atmospheric low-frequency variability or ocean circulation) and the different physical schemes of models on the sea ice trends. We find that the observed SIE trend falls within the range of the CMIP5 and CMIP6 distributions. If we assume that the trends are subject to a Gaussian distribution, there would be approximately 42% of the runs in CMIP6, versus 28% in CMIP5 that simulate a September Arctic SIE retreat rate faster than the observations. This result is consistent with the findings of Stroeve et al. (2012a), who suggested that around 25% of CMIP5 simulations display a September SIE trend that is as strong as, or stronger, than

FIG. 11. Scatterplot of the atmospheric model component resolution (a) and sea ice/ocean model component resolution (b) vs September SIE trend.

the observations during the satellite era. Under this assumption, the center of the distribution represents the SIE response to external forcing and the width of the distribution reflects the possible effect of internal variability on the SIE trend on decadal time scales. We can infer from the results that external forcing can explain approximately 77% (the mean of CMIP5 and CMIP6) of the observed September SIE decline. It should be noted, however, that all the results discussed above are based on the assumption that the simulations of these models conform to the normal distribution. In fact, they are not expected to be exactly Gaussian; indeed, as we can see from Fig. 1, the mean value is slightly greater than the median value, which indicates they are likely to be the positive skewness. In general, given the limited

sample size and model bias (e.g., the overestimation of the response to global scale changes in external radiative forcings), these results should be considered preliminary and viewed as an upper-bound estimate of the contribution of anthropogenic forcing.

7. Discussion and conclusions

In this study, we examine how well CMIP6 climate models simulate the climatological state, internal variability, and the linear trend of the Arctic sea ice cover and compare these results to those from their CMIP5 counterparts. In terms of the climatological state of SIE, the performances of the CMIP5 and

CMIP6 MME means are very similar. Although the underestimation of SIE declining trends in CMIP5 models remains in CMIP6, the simulated CMIP6 MME mean SIE trend is more consistent with the observations. It should be noted, however, that whether this progress in CMIP6 is due to the improvement in model physics or changes in external forcing is open to question. Compared with CMIP5 models, the CMIP6 models have smaller MME interquartile ranges in SIE climatological values and trends. However, the bias is larger in the CMIP6 models than CMIP5 models in terms of SIE internal variability.

In addition, we examine how well each individual CMIP6 model and their CMIP5 counterpart can reproduce the historical sea ice cover. To quantify a model's ability to reproduce multiple basic features of sea ice cover, a synthesis of the sea ice cover performance score for each individual CMIP5 and CMIP6 model is provided. Individual model performance varies greatly for different evaluation metrics. Some models simulate the climatological state of sea ice cover very well but cannot successfully reproduce the declining trends in sea ice, and vice versa. No individual model could emerge as “the best” overall. Although Notz (2015) pointed out that climate models are only tools and are not designed to replicate observations, their ability to reproduce the historical state can nevertheless give certain confidence in the projection of the future evolution of sea ice cover. Therefore, our sea ice error score can be used as an approach to reduce the spread in modeled sea ice projections by giving higher weights to those models with more realistic simulations of historical sea ice cover.

Importantly, understanding the possible causes of the bias might help with improving the physical realism of models. Different initial conditions might play a partial role in the establishment of modeled SIE bias, but other factors are also crucial. Here, from Fig. 11, we can see that the estimation of the September SIE trends varies remarkably among models belonging to the same family. For example, among the NCAR model family, two CESM2 models use the same land, sea ice, and ocean model components, differing only in their atmospheric models. Specifically, one uses the Community Atmosphere Model, version 6 (CESM2-CAM6), and the other uses the Whole Atmosphere Community Climate Model (CESM2-WACCM)—the high-top counterpart that spans a range of altitudes from Earth's surface to the thermosphere (Danabasoglu et al. 2020). Relative to CESM2, the use of WACCM in CESM2-WACCM narrows the bias in the September SIE trends. This may support previous results that a more comprehensive atmosphere model with a well-resolved stratosphere representation tends to better represent the processes associated with troposphere–stratosphere coupling and the feedback between sea ice and the atmosphere (Notz et al. 2013).

DeWeaver and Bitz (2006) suggested that the atmospheric model resolution can significantly influence the simulations of Arctic sea ice and surface wind changes. As seen in Fig. 11, a higher resolution in the oceanic and atmospheric components in CNRM-CM6.1-HR reduces the bias of the September SIE trend compared with that for CNRM-CM6.1. The same conclusion can also be reached by comparing MPI-ESM1.2-HR and MPI-ESM1.2-LR. However, no significant linear relationship is discernible between the September SIE trends and resolution across the full suite of models (Fig. 11), mainly because the discrepancies between modeled and observed

trends can be caused by a number of factors, such as inaccurate values of external forcing, an insufficient number of ensemble members, a model's inappropriate response to imposed external forcing, biases in sea ice thickness, as well as observational errors (Gregory et al. 2002; Mahlstein and Knutti 2012; Massonnet et al. 2018; Stroeve and Notz 2015; Winton 2011). The influence of any individual factor could be obscured by the complexity of the 36 CMIP6 models with their different model physics. Thus, to verify the importance of the resolution or any other factor, additional coordinated experiments should be applied.

To improve robustness of Arctic sea ice projections, estimation of the relative contributions of the internal variability of the climate system and increased anthropogenic emissions to the observed sea ice change is important. Based on the CMIP6 MME mean, we infer that approximately $22\% \pm 5\%$ of the September SIE decline from 1979 to 2014 can be attributed internal variability, versus $33\% \pm 3\%$ in their CMIP5 counterpart, which is consistent with the results reported by Stroeve et al. (2012a). This result is based on the assumption that the simulations of these models are subject to a normal distribution and that models have no bias in their response to external forcing. However, they are not expected to be exactly Gaussian, and some other researchers have pointed out a low sea ice sensitivity in some climate models (Mahlstein and Knutti 2012; Notz and Stroeve 2016; Rosenblum and Eisenman 2017). Thus, this value could be considered as a lower bound on the estimation of internal variability. In this regard, the CMIP6 projections with respect to the timing of an ice-free Arctic should be interpreted with caution.

Acknowledgments. We thank three anonymous reviewers and Editor James Screen for their constructive suggestions, which greatly helped to improve the quality of this manuscript. We acknowledge the World Climate Research Programme, which, through its Working Group, is responsible for CMIP. We thank the climate modeling centers for producing and making their model output available. We also thank the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant XDA19070404) and the National Natural Science Foundation of China (Grant 41725018).

Data availability statement. The sea ice concentration data from the United States National Snow and Ice Data Center (NSIDC) are available from <http://nsidc.org/data/seaice/>. The CMIP5 and CMIP6 data can be obtained from the ESGF nodes (<https://esgf-data.dkrz.de/projects/esgf-dkrz/>).

REFERENCES

- Årthun, M., T. Eldevik, L. Smedsrud, Ø. Skagseth, and R. Ingvaldsen, 2012: Quantifying the influence of Atlantic heat on Barents Sea ice variability and retreat. *J. Climate*, **25**, 4736–4743, <https://doi.org/10.1175/JCLI-D-11-00466.1>.
- Arzel, O., T. Fichefet, and H. Goosse, 2006: Sea ice evolution over the 20th and 21st centuries as simulated by current AOGCMs.

- Ocean Modell.*, **12**, 401–415, <https://doi.org/10.1016/j.ocemod.2005.08.002>.
- Bellenger, H., É. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard, 2014: ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dyn.*, **42**, 1999–2018, <https://doi.org/10.1007/s00382-013-1783-z>.
- Cavalieri, D. J., and C. L. Parkinson, 2012: Arctic sea ice variability and trends, 1979–2010. *Cryosphere*, **6**, 881–889, <https://doi.org/10.5194/tc-6-881-2012>.
- , —, P. Gloersen, and H. J. Zwally, 1996: Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, version 1 (updated yearly). NASA National Snow and Ice Data Center Distributed Active Archive Center, accessed 31 March 2020, <https://doi.org/10.5067/8GQ8LZQVLOVL>.
- Checa-Garcia, R., M. I. Hegglin, D. Kinnison, D. A. Plummer, and K. P. Shine, 2018: Historical tropospheric and stratospheric ozone radiative forcing using the CMIP6 database. *Geophys. Res. Lett.*, **45**, 3264–3273, <https://doi.org/10.1002/2017GL076770>.
- Comiso, J. C., 2012: Large decadal decline of the Arctic multiyear ice cover. *J. Climate*, **25**, 1176–1193, <https://doi.org/10.1175/JCLI-D-11-00113.1>.
- , 2017: Bootstrap sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS, version 3. NASA National Snow and Ice Data Center Distributed Active Archive Center, accessed 7 July 2020, <https://doi.org/10.5067/7Q8HCCWS4I0R>.
- Danabasoglu, G., and Coauthors, 2020: The Community Earth System Model version 2 (CESM2). *J. Adv. Model. Earth Syst.*, **12**, e2019MS001916, <https://doi.org/10.1029/2019MS001916>.
- DeRepentigny, P., A. Jahn, M. M. Holland, and A. Smith, 2020: Arctic sea ice in two configurations of the CESM2 during the 20th and 21st centuries. *J. Geophys. Res. Oceans*, **125**, e2020JC016133, <https://doi.org/10.1029/2020JC016133>.
- Deser, C., L. Sun, R. A. Tomas, and J. Screen, 2016: Does ocean coupling matter for the northern extratropical response to projected Arctic sea ice loss? *Geophys. Res. Lett.*, **43**, 2149–2157, <https://doi.org/10.1002/2016GL067792>.
- DeWeaver, E., and C. M. Bitz, 2006: Atmospheric circulation and its effect on Arctic sea ice in CCSM3 simulations at medium and high resolution. *J. Climate*, **19**, 2415–2436, <https://doi.org/10.1175/JCLI3753.1>.
- Ding, Q., J. M. Wallace, D. S. Battisti, E. J. Steig, A. J. Gallant, H.-J. Kim, and L. Geng, 2014: Tropical forcing of the recent rapid Arctic warming in northeastern Canada and Greenland. *Nature*, **509**, 209–212, <https://doi.org/10.1038/nature13260>.
- , 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>.
- Döscher, R., T. Vihma, and E. Maksimovich, 2014: Recent advances in understanding the Arctic climate system state and change from a sea ice perspective: A review. *Atmos. Chem. Phys.*, **14**, 13 571–13 600, <https://doi.org/10.5194/acp-14-13571-2014>.
- 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>.
- Fauria, M. M., A. Grinsted, S. Helama, J. Moore, M. Timonen, T. Martma, E. Isaksson, and M. Eronen, 2010: Unprecedented low twentieth century winter sea ice extent in the western Nordic Seas since AD 1200. *Climate Dyn.*, **34**, 781–795, <https://doi.org/10.1007/s00382-009-0610-z>.
- Fetterer, F., K. Knowles, W. N. Meier, M. Savoie, and A. K. Windnagel, 2017: Sea ice index, version 3 (updated daily). National Snow and Ice Data Center, accessed 27 March 2020, <https://doi.org/10.7265/N5K072F8>.
- Francis, J. A., W. Chan, D. J. Leathers, J. R. Miller, and D. E. Veron, 2009: Winter Northern Hemisphere weather patterns remember summer Arctic sea-ice extent. *Geophys. Res. Lett.*, **36**, L07503, <https://doi.org/10.1029/2009GL037274>.
- Gidden, M., and Coauthors, 2019: Global emissions pathways under different socioeconomic scenarios for use in CMIP6: A dataset of harmonized emissions trajectories through the end of the century. *Geosci. Model Dev.*, **12**, 1443–1475, <https://doi.org/10.5194/gmd-12-1443-2019>.
- Giorgi, F., and L. O. Mearns, 2002: Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the “reliability ensemble averaging” (REA) method. *J. Climate*, **15**, 1141–1158, [https://doi.org/10.1175/1520-0442\(2002\)015<1141:COAURA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1141:COAURA>2.0.CO;2).
- Gregory, J. M., P. Stott, D. Cresswell, N. Rayner, C. Gordon, and D. 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>.
- Holland, M. M., D. A. Bailey, B. P. Briegleb, B. Light, and E. Hunke, 2012: Improved sea ice shortwave radiation physics in CCSM4: The impact of melt ponds and aerosols on Arctic sea ice. *J. Climate*, **25**, 1413–1430, <https://doi.org/10.1175/JCLI-D-11-00078.1>.
- IPCC, 2013: *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, 1535 pp., <https://doi.org/10.1017/CBO9781107415324>.
- Jahn, A., 2018: Reduced probability of ice-free summers for 1.5°C compared to 2°C warming. *Nat. Climate Change*, **8**, 409–413, <https://doi.org/10.1038/s41558-018-0127-8>.
- 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>.
- Kinnard, C., C. M. Zdanowicz, D. A. Fisher, E. Isaksson, A. de Vernal, and L. G. Thompson, 2011: Reconstructed changes in Arctic sea ice over the past 1,450 years. *Nature*, **479**, 509–512, <https://doi.org/10.1038/nature10581>.
- Koldunov, N. V., D. Stammer, and J. Marotzke, 2010: Present-day Arctic sea ice variability in the coupled ECHAM5/MPI-OM model. *J. Climate*, **23**, 2520–2543, <https://doi.org/10.1175/2009JCLI3065.1>.
- 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, J., M. Song, R. M. Horton, and Y. Hu, 2013: Reducing spread in climate model projections of a September ice-free Arctic. *Proc. Natl. Acad. Sci. USA*, **110**, 12 571–12 576, <https://doi.org/10.1073/pnas.1219716110>.
- Mahlstein, I., and R. Knutti, 2012: September Arctic sea ice predicted to disappear near 2°C global warming above present. *J. Geophys. Res.*, **117**, <https://doi.org/10.1029/2011JD016709>.
- Maslanik, J., C. Fowler, J. Stroeve, S. Drobot, J. Zwally, D. Yi, and W. Emery, 2007: A younger, thinner Arctic ice cover: Increased potential for rapid, extensive sea-ice loss. *Geophys. Res. Lett.*, **34**, L24501, <https://doi.org/10.1029/2007GL032043>.
- Massonnet, F., M. Vancoppenolle, H. Goosse, D. Docquier, T. Fichet, 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>.
- Meehl, G. A., and Coauthors, 2012: Climate system response to external forcings and climate change projections in CCSM4. *J. Climate*, **25**, 3661–3683, <https://doi.org/10.1175/JCLI-D-11-00240.1>.

- Meier, W., F. Fetterer, M. Savoie, S. Mallory, R. Duerr, and J. Stroeve, 2017: NOAA/NSIDC climate data record of passive microwave sea ice concentration, version 3. National Snow and Ice Data Center, accessed 7 July 2020, <https://doi.org/10.7265/N59P2ZTG>.
- Melsom, A., V. S. Lien, and W. P. Budgell, 2009: Using the Regional Ocean Modeling System (ROMS) to improve the ocean circulation from a GCM 20th century simulation. *Ocean Dyn.*, **59**, 969–981, <https://doi.org/10.1007/s10236-009-0222-5>.
- Mioduszewski, J. R., S. Vavrus, M. Wang, M. Holland, and L. Landrum, 2019: Past and future interannual variability in Arctic sea ice in coupled climate models. *Cryosphere*, **13**, 113–124, <https://doi.org/10.5194/tc-13-113-2019>.
- Notz, D., 2014: Sea-ice extent and its trend provide limited metrics of model performance. *Cryosphere*, **8**, 229–243, <https://doi.org/10.5194/tc-8-229-2014>.
- , 2015: How well must climate models agree with observations? *Philos. Trans. Roy. Soc.*, **373A**, 20140164, do, <https://doi.org/10.1098/rsta.2014.0164>.
- , and J. Stroeve, 2016: Observed Arctic sea-ice loss directly follows anthropogenic CO₂ emission. *Science*, **354**, 747–750, <https://doi.org/10.1126/science.aag2345>.
- , F. A. Haumann, H. Haak, J. H. Jungclauss, and J. Marotzke, 2013: Arctic sea-ice evolution as modeled by Max Planck Institute for Meteorology's Earth system model. *J. Adv. Model. Earth Syst.*, **5**, 173–194, <https://doi.org/10.1002/jame.20016>.
- , H., and Coauthors, 2020: Arctic sea ice in CMIP6. *Geophys. Res. Lett.*, **47**, e2019GL086749, <https://doi.org/10.1029/2019GL086749>.
- Overland, J. E., and M. Wang, 2007: Future regional Arctic sea ice declines. *Geophys. Res. Lett.*, **34**, L17705, <https://doi.org/10.1029/2007GL030808>.
- , and —, 2013: When will the summer Arctic be nearly sea ice free? *Geophys. Res. Lett.*, **40**, 2097–2101, <https://doi.org/10.1002/grl.50316>.
- Parkinson, C. L., and J. C. Comiso, 2013: On the 2012 record low Arctic sea ice cover: Combined impact of preconditioning and an August storm. *Geophys. Res. Lett.*, **40**, 1356–1361, <https://doi.org/10.1002/grl.50349>.
- , D. J. Cavalieri, P. Gloersen, H. J. Zwally, and J. C. Comiso, 1999: Arctic sea ice extents, areas, and trends, 1978–1996. *J. Geophys. Res.*, **104**, 20 837–20 856, <https://doi.org/10.1029/1999JC900082>.
- Pedersen, C. A., E. Roeckner, M. Luthje, and J.-G. Winther, 2009: A new sea ice albedo scheme including melt ponds for ECHAM5 general circulation model. *J. Geophys. Res.*, **114**, D08101, <https://doi.org/10.1029/2008JD010440>.
- Rampal, P., J. Weiss, C. Dubois, and J.-M. Campin, 2011: IPCC climate models do not capture Arctic sea ice drift acceleration: Consequences in terms of projected sea ice thinning and decline. *J. Geophys. Res.*, **116**, C00D07, <https://doi.org/10.1029/2011JC007110>.
- Roeckner, E., T. Mauritsen, M. Esch, and R. Brokopf, 2012: Impact of melt ponds on Arctic sea ice in past and future climates as simulated by MPI-ESM. *J. Adv. Model. Earth Syst.*, **4**, M00A02, <https://doi.org/10.1029/2012MS000157>.
- 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>.
- Sanderson, B. M., R. Knutti, and P. Caldwell, 2015: A representative democracy to reduce interdependency in a multimodel ensemble. *J. Climate*, **28**, 5171–5194, <https://doi.org/10.1175/JCLI-D-14-00362.1>.
- , M. Wehner, and R. Knutti, 2017: Skill and independence weighting for multi-model assessments. *Geosci. Model Dev.*, **10**, 2379–2395, <https://doi.org/10.5194/gmd-10-2379-2017>.
- Santer, B. D., and Coauthors, 2008: Consistency of modelled and observed temperature trends in the tropical troposphere. *Int. J. Climatol.*, **28**, 1703–1722, <https://doi.org/10.1002/joc.1756>.
- Screen, J. A., 2018: Arctic sea ice at 1.5 and 2°C. *Nat. Climate Change*, **8**, 362–363, <https://doi.org/10.1038/s41558-018-0137-6>.
- , and I. Simmonds, 2010: The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature*, **464**, 1334–1337, <https://doi.org/10.1038/nature09051>.
- , and J. A. Francis, 2016: Contribution of sea-ice loss to Arctic amplification is regulated by Pacific Ocean decadal variability. *Nat. Climate Change*, **6**, 856–860, <https://doi.org/10.1038/nclimate3011>.
- , and R. Blackport, 2019: How robust is the atmospheric response to projected Arctic sea ice loss across climate models? *Geophys. Res. Lett.*, **46**, 11 406–11 415, <https://doi.org/10.1029/2019GL084936>.
- Serreze, M. C., and R. G. Barry, 2011: Processes and impacts of Arctic amplification: A research synthesis. *Global Planet. Change*, **77**, 85–96, <https://doi.org/10.1016/j.gloplacha.2011.03.004>.
- Smedsrud, L. H., and Coauthors, 2013: The role of the Barents Sea in the Arctic climate system. *Rev. Geophys.*, **51**, 415–449, <https://doi.org/10.1002/rog.20017>.
- Stroeve, J. C., and D. Notz, 2015: Insights on past and future sea-ice evolution from combining observations and models. *Global Planet. Change*, **135**, 119–132, <https://doi.org/10.1016/j.gloplacha.2015.10.011>.
- , M. M. Holland, W. Meier, T. Scambos, and M. Serreze, 2007: Arctic sea ice decline: Faster than forecast. *Geophys. Res. Lett.*, **34**, L09501, <https://doi.org/10.1029/2007GL029703>.
- , V. Kattsov, A. Barrett, M. Serreze, T. Pavlova, M. Holland, and W. N. Meier, 2012a: Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations. *Geophys. Res. Lett.*, **39**, L16502, <https://doi.org/10.1029/2012GL052676>.
- , M. C. Serreze, M. M. Holland, J. E. Kay, J. Malanik, and A. P. Barrett, 2012b: The Arctic's rapidly shrinking sea ice cover: A research synthesis. *Climatic Change*, **110**, 1005–1027, <https://doi.org/10.1007/s10584-011-0101-1>.
- Thompson, D. W., E. A. Barnes, C. Deser, W. E. Foust, and A. S. Phillips, 2015: Quantifying the role of internal climate variability in future climate trends. *J. Climate*, **28**, 6443–6456, <https://doi.org/10.1175/JCLI-D-14-00830.1>.
- Wang, M., and J. E. Overland, 2012: A sea ice free summer Arctic within 30 years: An update from CMIP5 models. *Geophys. Res. Lett.*, **39**, L18501, <https://doi.org/10.1029/2012GL052868>.
- 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>.
- Zhang, X., 2010: Sensitivity of Arctic summer sea ice coverage to global warming forcing: Towards reducing uncertainty in Arctic climate change projections. *Tellus*, **62A**, 220–227, <https://doi.org/10.1111/j.1600-0870.2010.00441.x>.