Arctic Warming Revealed by Multiple CMIP6 Models: Evaluation of Historical Simulations and Quantification of Future Projection Uncertainties

ZIYI CAI,[a] QINGLONG YOU,[a,b] FANGYING WU,[c] HANS W. CHEN,[d] DELIANG CHEN,[e] AND JUDAH COHEN[f,g]

[a] *Department of Atmospheric and Oceanic Sciences, Institute of Atmospheric Sciences, Fudan University, Shanghai, China*
[b] *Innovation Center of Ocean and Atmosphere System, Zhuhai Fudan Innovation Research Institute, Zhuhai, China*
[c] *Key Laboratory of Meteorological Disaster, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China*
[d] *Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden*
[e] *Regional Climate Group, Department of Earth Sciences, University of Gothenburg, Gothenburg, Sweden*
[f] *Atmospheric and Environmental Research Inc., Lexington, Massachusetts*
[g] *Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Manuscript received 9 October 2020, in final form 11 March 2021)

ABSTRACT: The Arctic has experienced a warming rate higher than the global mean in the past decades, but previous studies show that there are large uncertainties associated with future Arctic temperature projections. In this study, near-surface mean temperatures in the Arctic are analyzed from 22 models participating in phase 6 of the Coupled Model Intercomparison Project (CMIP6). Compared with the ERA5 reanalysis, most CMIP6 models underestimate the observed mean temperature in the Arctic during 1979–2014. The largest cold biases are found over the Greenland Sea the Barents Sea, and the Kara Sea. Under the SSP1-2.6, SSP2-4.5, and SSP5-8.5 scenarios, the multimodel ensemble mean of 22 CMIP6 models exhibits significant Arctic warming in the future and the warming rate is more than twice that of the global/Northern Hemisphere mean. Model spread is the largest contributor to the overall uncertainty in projections, which accounts for 55.4% of the total uncertainty at the start of projections in 2015 and remains at 32.9% at the end of projections in 2095. Internal variability uncertainty accounts for 39.3% of the total uncertainty at the start of projections but decreases to 6.5% at the end of the twenty-first century, while scenario uncertainty rapidly increases from 5.3% to 60.7% over the period from 2015 to 2095. It is found that the largest model uncertainties are consistent cold bias in the oceanic regions in the models, which is connected with excessive sea ice area caused by the weak Atlantic poleward heat transport. These results suggest that large intermodel spread and uncertainties exist in the CMIP6 models' simulation and projection of the Arctic near-surface temperature and that there are different responses over the ocean and land in the Arctic to greenhouse gas forcing. Future research needs to pay more attention to the different characteristics and mechanisms of Arctic Ocean and land warming to reduce the spread.

KEYWORDS: Arctic; Climate prediction; Temperature; Coupled models; Model evaluation/performance

1. Introduction

Enhanced warming in the Arctic (north of 66°N) is seen in recent observations and model simulations, a phenomenon called Arctic amplification (AA) (Cohen et al. 2014; Screen and Simmonds 2010). AA acts as a robust response to global climate change and could have profound impacts on the climate system, resulting in Arctic sea ice shrinkage (Box et al. 2019; Richter-Menge et al. 2020; Stroeve and Notz 2018), diminished mass balance of the Greenland ice sheet (Mouginot et al. 2019; Pattyn et al. 2018), and release of the potent greenhouse gas methane from the thawing Arctic permafrost (Portnov et al. 2016). Additionally, there are indications that AA can intensify extreme climate events in the northern midlatitudes, but this view remains controversial (Blackport and Screen 2020; Cohen et al. 2014, 2020; Deng et al. 2020). Thus, it is essential to understand climate processes in the Arctic to improve our understanding of future climate change. A major challenge, however, is the limited observations and in situ measurements in the polar region (Cowtan and Way 2014). Therefore, some studies are based on interpolation and other statistical methods to fill data gaps in the Arctic (Dodd et al. 2015; Huang et al. 2017; Rohde et al. 2013).

As numerical models have improved, climate models have become one of the primary tools for understanding patterns and mechanisms of climate change and for projecting future climate. However, climate feedbacks caused by the complex interactions between the ocean, sea ice, and atmosphere make simulating and projecting Arctic warming particularly challenging, and large uncertainties remain (Knutti 2008; Overland et al. 2011; Serreze and Francis 2006). Models cannot well represent local feedbacks that have great influences on Arctic warming such as sea ice-albedo feedback, Planck feedback, lapse-rate feedback, and cloud and water vapor feedback (Bonan et al. 2018; Goosse et al. 2018; Hu et al. 2020; Pithan and Mauritsen 2014; Stuecker et al. 2018). For example, models lack consistency in simulating the magnitude and spatial pattern of sea ice loss, new open water, and troposphere–stratosphere coupling (Screen et al. 2018), which will contribute to the intermodel spread in ocean–atmosphere heat exchange (Boeke and Taylor 2018) and atmosphere response to Arctic sea ice loss (Screen et al. 2018). Additionally, many models simulate too little liquid water in

Corresponding author: Prof. Qinglong You, qlyou@fudan.edu.cn; yqingl@126.com

DOI: 10.1175/JCLI-D-20-0791.1

TABLE 1. Information about the CMIP6 models in this study, including model name, originating group/country, horizontal resolution for the atmospheric component, and data availability for this study (indicated with a check mark ✓).

	Model	Country	Horizontal resolution	Surface temperature	Sea ice area percentage	Northward ocean heat transport
1	ACCESS-CM2	Australia	192 × 144	✓	✓	—
2	AWI-CM-1-1-MR	Germany	384 × 192	✓	—	—
3	BCC-CSM2-MR	China	320 × 160	✓	✓	—
4	CAMS-CSM1-0	China	320 × 160	✓	✓	—
5	CanESM5	Canada	128 × 64	✓	✓	✓
6	CESM2	United States	288 × 192	✓	✓	—
7	CESM2-WACCM	United States	288 × 192	✓	✓	—
8	EC-Earth3	European	512 × 256	✓	✓	✓
9	EC-Earth3-Veg	European	512 × 256	✓	✓	✓
10	FGOALS-f3-L	China	288 × 180	✓	✓	✓
11	FGOALS-g3	China	188 × 80	✓	✓	✓
12	FIO-ESM2-0	China	288 × 192	✓	✓	—
13	GFDL-ESM4	United States	288 × 180	✓	✓	—
14	INM-CM4-8	Russia	180 × 120	✓	✓	—
15	INM-CM5-0	Russia	180 × 120	✓	✓	—
16	IPSL-CM6A-LR	France	144 × 143	✓	✓	✓
17	MIROC6	Japan	256 × 128	✓	✓	✓
18	MPI-ESM1-2-HR	Germany	384 × 192	✓	✓	✓
19	MPI-ESM1-2-LR	Germany	192 × 96	✓	✓	✓
20	MRI-ESM2-0	Japan	320 × 160	✓	✓	✓
21	NESM3	China	192 × 96	✓	✓	—
22	NorESM2-LM	Norway	144 × 96	✓	✓	✓

low-level Arctic clouds, which leads to errors in surface downwelling radiation and surface temperature response (Barton et al. 2014; Karlsson and Svensson 2013). Nonlocal processes cannot be ignored either (Bonan et al. 2018), as differences in the northward ocean heat flux among models lead to intermodel spread of sea ice concentration, which affects the surface temperature through the energy balance budget (Hodson et al. 2013; Mahlstein and Knutti 2011). Understanding the different responses of global climate models (GCMs) in reproducing Arctic temperature and the reasons for the simulation biases constitutes the basis for projecting future Arctic warming.

The simulations and projections of climate change in the Arctic have been assessed using GCMs from phases 3 and 5 of the Coupled Model Intercomparison Project (CMIP3 and CMIP5) under different scenarios of greenhouse gas emissions (Chapman and Walsh 2007; Hao et al. 2018; Liu et al. 2008; Tao et al. 1996). For example, many previous studies have focused on the performance of GCMs in reproducing the Arctic sea ice extent and thickness and surface radiation budget, which are essential to understand the ''new Arctic'' in the future climate (Barton et al. 2014; Davy and Outten 2020; Landrum and Holland 2020; Senftleben et al. 2020; Wu et al. 2019). In terms of surface mean temperature, the Fourth and Fifth Assessment Reports of the Intergovernmental Panel on Climate Change (IPCC AR4 and AR5) used the CMIP3/CMIP5 GCMs to evaluate the intermodel differences of global mean temperature (IPCC 2007, 2013). Some studies suggest that a large variation of intermodel differences exists in modeling the magnitude of Arctic warming, and most GCMs have large cold biases, while the temperature trend is overestimated (Chapman and Walsh 2007; Davy and Outten 2020; Hao et al. 2018; Huang et al. 2019; Koenigk et al. 2013; Wang et al. 2020; Wei et al. 2018; Xie et al. 2016; X. Zhou et al. 2019).

Compared with CMIP5 models, the new-generation CMIP6 models have higher resolutions and are run using a combination of a representative concentration pathway (RCP) and a Shared Socioeconomic Pathway (SSP) scenario (Eyring et al. 2016; Taylor et al. 2012). Although a previous study has evaluated Arctic near-surface temperatures in the CMIP6 models (Davy and Outten 2020; You et al. 2021), there have only been limited studies on quantifying the uncertainty in the CMIP6 Arctic-only projections. Quantitative research on uncertainty can help to understand the deficiencies in the models' abilities to accurately simulate Arctic temperatures. Thus, this study aims to 1) evaluate the performance of the CMIP6 models in simulating Arctic near-surface temperatures during the contemporary period (1979–2014); 2) assess future climate change in the Arctic based on the CMIP6 model simulations in the near-term (2021–40), midterm (2041–60), and long-term (2081–2100); and 3) quantitatively analyze the uncertainties in the model projections.

2. Data and methods

a. Data

The monthly near-surface mean temperature outputs from the historical experiment and Scenario Model Intercomparison Project (ScenarioMIP) of the 22 CMIP6 models were collected (Table 1). The monthly sea ice area concentration and northward ocean heat transport output from the historical experiment were used in this study (Table 1). The model outputs are

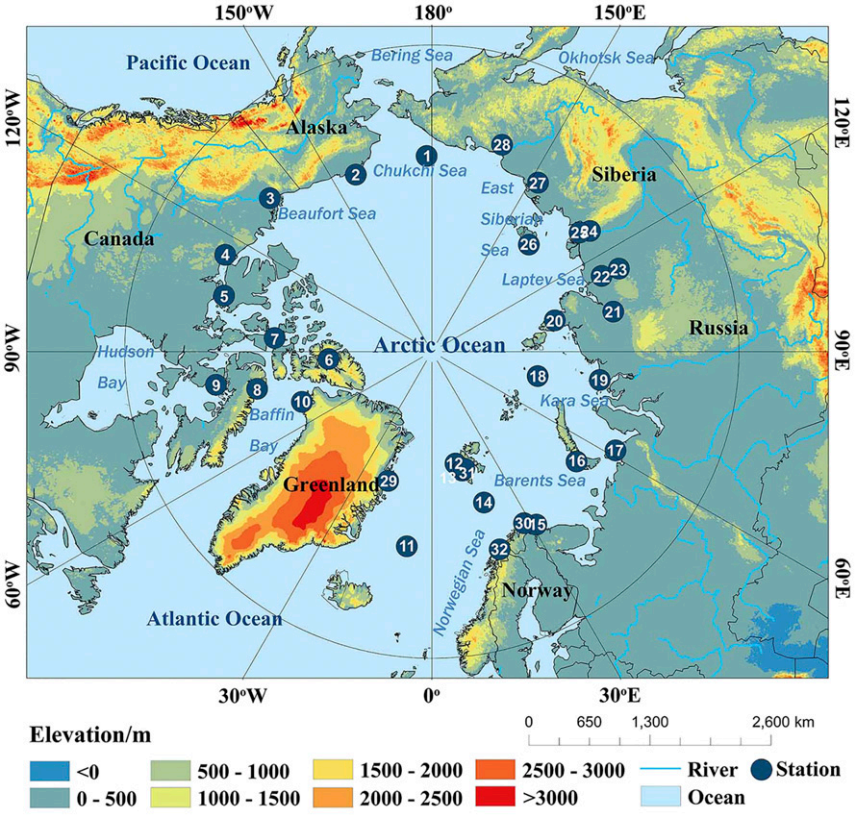

FIG. 1. Map of Arctic land surface elevation and location of meteorological stations used in this study. The numbers 1–32 refer to the stations listed in Table 2.

available from https://esgf-node.llnl.gov/search/cmip6/. Because many models lack a sufficiently large ensemble at this time, only the first ensemble member of each model (r1i1p1f1) is selected to make the comparisons fair. The future scenarios are based on three Shared Socioeconomic Pathways (SSPs): SSP1-2.6, SSP2-4.5, and SSP5-8.5. These three scenarios add the SSP1, SSP2, and SSP5 socioeconomic pathways to the RCP2.6, RCP4.5, and RCP8.5 radiative forcing scenarios, respectively (Gidden et al. 2019; O'Neill et al. 2016).

Most areas of the Arctic are covered by the Arctic Ocean, and the observational conditions are relatively harsh (Fig. 1). Because there are almost no meteorological stations above 75°N in the Arctic, the monthly mean 2-m temperature data from the ERA5 reanalysis (Hersbach et al. 2020) are used as the benchmark field to evaluate the CMIP6 models.

Although ERA5 is constrained by observations, the reanalysis could still suffer from biases from the underlying model (Graham et al. 2019; Lindsay et al. 2014; Wang et al. 2019). Therefore, we also evaluated the 2-m temperatures in ERA5 against observations from meteorological stations and the Berkeley Earth Surface Temperature (BEST) dataset in the Arctic during the period 1979–2014. The surface air temperature station observations were taken from the National Aeronautics and Space Administration's (NASA) Goddard Institute for Space Studies surface temperature analysis v4 (GISTEMP) (Lenssen et al. 2019; Rohde and Hausfather 2020), which is available from https://data.giss.nasa.gov/gistemp/station_data_v4_globe/. Considering the completeness of the time series and the location of the stations, 32 stations within the Arctic Circle were selected and compared with the data in the closest grid points in ERA5. The information of the 32 stations is shown in Table 2. The BEST dataset is available from http://berkeleyearth.org/data/ (Rohde et al. 2013; Rohde and Hausfather 2020).

b. *Model performance evaluation analysis*

The CMIP6 model simulations were evaluated against the ERA5 benchmark by first regridding the model and reanalysis fields onto a common 1.5° × 1.5° grid using bilinear interpolation and calculating the annual mean temperatures based on the monthly data. To evaluate the spatial characteristics, we temporally averaged the near-surface temperature field during 1979–2014 to minimize the influence of internal variability. Then, we calculated the correlation coefficient (R), centered root-mean-square error (RMSE), and ratio of standard deviations (RSTD) between the simulation and benchmark, and visualized these metrics using Taylor diagrams (Taylor 2001). The best model simulation results for spatial characteristics are when the R and RSTD are close to 1 and RMSE is close to 0. Note that the centered RMSEs do not account for overall biases in the models. Model biases are instead reflected in the temporal evaluation metrics. For evaluating the temporal

TABLE 2. Observational station information: name, coordinate, elevation, temperature trend (1979–2014), mean temperature (1979–2014). All trends are significant at the 0.05 significance level.

	Station	Longitude	Latitude	Elevation (m)	Temperature trend (°C decade^{-1})	Mean temperature (°C)
1	Ostrov Vrangelja	178.4833°W	70.9831°N	2.0	0.85	−10.1
2	Barrow Post Rogers Ap	156.7814°W	71.2833°N	9.4	0.37	−10.2
3	Inuvik A	133.4833°W	68.3000°N	68.0	0.77	−7.9
4	Kugluktuk A	115.1500°W	67.8167°N	23.0	0.61	−10.2
5	Cambridge Bay A	105.1333°W	69.1000°N	31.0	0.63	−13.8
6	Eureka	85.9333°W	79.9833°N	10.0	0.81	−18.8
7	Resolute Cs	94.9833°W	74.7167°N	30.0	0.78	−15.6
8	Pond Inlet	77.9667°W	72.6833°N	62.0	0.69	−14.1
9	Hall Beach A	81.2500°W	68.7833°N	9.0	0.80	−13.3
10	Thule Op Site	68.8333°W	76.5167°N	77.1	0.72	−10.8
11	Jan Mayen	7.3300°W	70.9300°N	10.0	0.75	−0.2
12	Ny Alesund	11.9331°E	78.9230°N	8.0	0.91	−5.1
13	Barencburg	14.3000°E	78.1000°N	49.0	1.03	−4.7
14	Bjoernoeya	19.0167°E	74.5167°N	16.0	0.81	−1.1
15	Vardo	31.1000°E	70.3670°N	15.0	0.61	2.1
16	Maliye Karmakuly	52.7300°E	72.3794°N	18.0	0.79	−4.4
17	Amderma	61.7000°E	69.7500°N	53.0	0.69	−6.1
18	Ostrov Vize	76.9800°E	79.5000°N	10.0	1.53	−13.0
19	Dikson	80.4000°E	73.5000°N	42.0	0.78	−10.8
20	Gmo Imekfedorova	104.3000°E	77.7200°N	12.0	1.10	−14.2
21	Hatanga	102.4667°E	71.9831°N	31.0	0.68	−12.3
22	Saskylah	114.0800°E	71.9700°N	16.0	0.83	−13.8
23	Dzalinda 1	113.9700°E	70.1297°N	61.0	0.86	−13.0
24	Kjusjur	127.4000°E	70.6800°N	30.0	0.59	−13.1
25	Tiksi	128.9197°E	71.5800°N	6.0	0.64	−12.7
26	Ostrov Kotelnyj	137.8700°E	76.0000°N	12.0	0.97	−14.1
27	Chokurdah	147.8831°E	70.6167°N	44.0	0.72	−13.1
28	Cherskij	161.2830°E	68.7500°N	28.0	0.77	−10.4
29	Danmarkshavn	18.6700°W	76.7700°N	12.0	0.50	−11.4
30	Slettnes Fyr	28.2178°E	71.0839°N	8.0	0.39	2.2
31	Svalbard Airport	15.4667°E	78.2500°N	28.0	1.19	−4.9
32	Tromsolangnes	18.9131°E	69.6767°N	8.0	0.53	2.8

temperature characteristics in the Arctic, we averaged the near-surface mean temperatures northward of 66°N and then compared the means, trends, and interannual variations of the annual mean temperature time series between the models and benchmark. The performance of the mean temperatures and the temperature trends is based on the absolute difference of the models and benchmark, and the smallest absolute bias represents the best performance. An interannual variability skill (IVS) was calculated to measure the accuracy of the model to simulate the interannual variations of the Arctic near-surface temperature (Chen et al. 2011):

$$\mathrm{IVS} = \left(\frac{\sigma_{to}}{\sigma_{tm}} - \frac{\sigma_{tm}}{\sigma_{to}}\right)^2,$$

where σ_{to} and σ_{tm} denote the interannual standard deviations of the benchmark and simulation, respectively. An IVS value closer to 0 indicates higher skill while a large IVS value indicates poor skill.

We also summarized the models' overall performances using a spatial skill score and a temporal rank score. The spatial skill score was calculated using the Taylor skill score (TS) (Taylor 2001):

$$\mathrm{TS} = \frac{4(1+R)^2}{\left(\frac{\sigma_{sm}}{\sigma_{so}} + \frac{\sigma_{so}}{\sigma_{sm}}\right)^2 (1+R_0)^2},$$

where σ_{sm} and σ_{so} are the standard deviations of the simulated and benchmark spatial fields, R is the spatial correlation coefficient for the evaluated model, and R_0 is the maximum spatial correlation coefficient over all models. Theoretically, a TS value of 1.0 indicates that the simulation exactly corresponds to the benchmark, while a smaller TS value indicates a poorer fit.

The temporal rank score (TR) was calculated based on the differences in Arctic near-surface mean temperatures, temperature trends, and interannual variations between the models and benchmark. The model with the smallest absolute difference of the mean temperature and temperature trend from the benchmark and the IVS closest to 0 performs best. The average of the model rankings of these three indicators is used as a TR of the model, with smaller values indicating higher skill.

Finally, we ranked the models according to their TS and TR values. To transform the model rankings rank$_i$ (ranging from 1

to m, where m is the number of models and rank 1 is the best performing model) to a number between 0 and 1, we calculated an M_R value (Jiang et al. 2015) for each model:

$$M_R = 1 - \frac{\text{rank}_i}{m}.$$

An M_R value closer to 1 indicates a better fit of the model to the benchmark.

c. Uncertainty analysis

Uncertainty in climate projections arises from three distinct sources (Hawkins and Sutton 2009, 2011; Lehner et al. 2020). The first uncertainty is due to the internal variability of the climate system, which represents the uncertainty in the state of the climate system in the absence of any anthropogenic radiative forcing of the planet. The second is the model uncertainty due to model structural differences, which causes different models to produce different projections for the same radiative forcing and internal variability. The third is scenario uncertainty due to uncertainties in the future radiative forcing and socioeconomic pathway (Hawkins and Sutton 2009, 2011; Lehner et al. 2020). Below we follow the methods proposed by Hawkins and Sutton (2009, 2011) to separate and quantify these uncertainty sources.

Each individual prediction is estimated using ordinary least squares with a fourth-order polynomial during the period 1854–2095 (9-yr running window over 1850–2099). The predictions X for each model m, scenario s, and year t can be written as

$$X_{m,s,t} = x_{m,s,t} + i_{m,s} + \varepsilon_{m,s,t},$$

where $x_{m,s,t}$ is the smooth fit, $i_{m,s}$ is the reference temperature (mean of years 1986–2005) that was estimated from the smooth fit, and $\varepsilon_{m,s,t}$ is the residual.

The internal variability component V is computed from the multimodel mean of the variances of the residuals:

$$V = \sum \frac{1}{n} \text{var}_{s,t}(\varepsilon_{m,s,t}),$$

where n is the number of models ($n = 22$) and $\text{var}_{s,t}$ denotes the variance across all scenarios and across time.

The model uncertainty for each scenario is estimated from the variance in the different model prediction estimate. The multiscenario mean of the variance of the model prediction fits is taken as an estimate of the model uncertainty $M(t)$:

$$M(t) = \frac{1}{N_s} \sum_s \text{var}_m(x_{m,s,t}),$$

where N_s is the number of scenarios ($N_s = 3$).

The scenario uncertainty $S(t)$ is calculated from the variance of the multimodel mean over the three scenarios (SSP1-2.6, SSP2-4.5, and SSP5-8.5):

$$S(t) = \text{var}_s\left(\sum_m \frac{1}{n} x_{m,s,t}\right).$$

It is assumed that the three uncertainty sources are all independent. Thus, the total variance $T(t)$ is

$$T(t) = V + S(t) + M(t),$$

and the mean change of all the predictions $G(t)$ above the reference temperature is

$$G(t) = \frac{1}{N_s} \sum_{m,s} \frac{1}{n} x_{m,s,t}.$$

The total fractional uncertainty $F(t)$ (90% confidence level) is calculated as follows:

$$F(t) = \frac{1.65\sqrt{T(t)}}{G(t)},$$

where 1.65 is the Z score for the 90% confidence interval. The corresponding fractional uncertainty for internal variability $F_V(t)$, model uncertainty $F_M(t)$, and scenario uncertainty $F_S(t)$ (90% confidence level) are calculated as $1.65\sqrt{V}/G(t)$, $1.65\sqrt{M(t)}/G(t)$, and $1.65\sqrt{S(t)}/G(t)$, respectively. The fraction of total variance is the proportion of each component to the total fractional uncertainty $F(t)$, which defines the relative importance of the three uncertainty sources. This method has been widely used in the uncertainty analysis of regional climate projections (Hawkins and Sutton 2009, 2011; Hodson et al. 2013; Wu et al. 2020). Lehner et al. (2020) reported that single model initial-condition large ensembles (SMILEs) allow a more robust separation between sources of uncertainty, especially in the estimation of internal variability. We also investigated the role of internal variability from six CMIP6 models (CanESM5, EC-Earth3, GISS-E2-1-G, IPSL-CM6A-LR, MIROC6, and ACCESS-ESM1-5) that already provided more than 10 ensembles at this time.

The signal-to-noise ratio (σ), which can also be used to assess the reliability of future projections, is usually defined as

$$\sigma = \frac{\bar{x}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}},$$

where n is the number of models ($n = 22$), $\bar{x}$ is the mean change of a projection, and x_i is the change of an individual model's projection. The numerator is taken as the signal and the denominator as the noise, with the latter calculated as the standard deviation between the models. When $\sigma > 1$ this indicates that the signal is greater than the noise, that is, the projected change is distinguishable from internal variability and model error. A larger σ value typically indicates a higher reliability in the forced response (Zhou and Yu 2006).

3. Results

a. Evaluation of CMIP6 models' performance

To measure the accuracy of the ERA5 reanalysis, we compared it with the ground-based station observations shown in Fig. 1 and Table 2. Figure 2 shows the time series of the annual near-surface mean temperature of the Arctic from 1979 to 2014 and the differences in near-surface mean temperatures and temperature trends between ERA5 and observations during

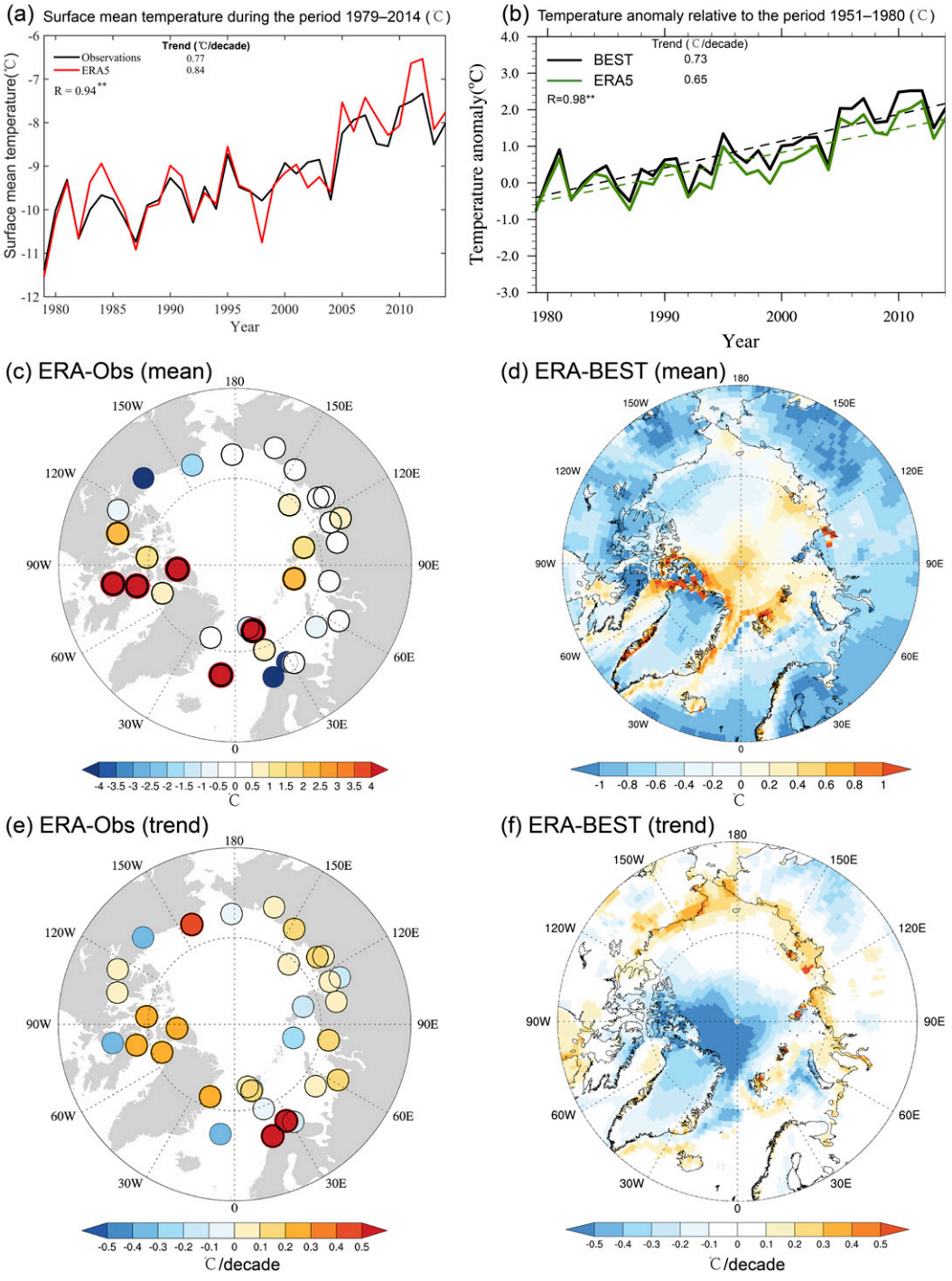

FIG. 2. Comparison of (left) the ERA5 reanalysis with 32 observational stations and (right) the Berkeley Earth Surface Temperatures dataset (BEST) during 1979–2014. (a) Time series of annual mean near-surface temperature of 32 observational stations and the corresponding closest grid point in ERA5. (b) Time series of annual mean near-surface temperature of BEST and ERA5 (relative to the period during 1951–80). Here we used the ERA-BE (back extension) to compute the 1951–80 baseline. (c) Differences in the near-surface mean temperature (°C) of the Arctic from 1979 to 2014 between the ERA5 reanalysis and observational stations (ERA5 minus observations). (d) As in (c), but for differences between ERA5 and BEST (ERA5 minus BEST). (e),(f) As in (c) and (d), but for differences in near-surface temperature trend (°C decade^{-1}).

1979–2014. The overall bias in the ERA5 reanalysis compared with the observed near-surface mean temperature of the Arctic during the period is 0.38°C, and the bias is larger in the Western Hemisphere than in the Eastern Hemisphere. The mean difference between the trends estimated from ERA5 and the observations is 0.11°C decade^{-1}, and the largest bias was found in northern Norway. The average correlation coefficient between the ERA5 reanalysis and observed time series reaches 0.94. Considering that the stations are only distributed over Arctic land, the BEST dataset was used to verify the performance of ERA5 over the Arctic Ocean (Fig. 2). The results show that ERA5 has a slight cold bias compared with the BEST dataset, and a slight warm bias over some parts of the ocean. The spatial distributions of the temperature trends in the two datasets are generally similar except for a small area north of Greenland, where the BEST and ERA5 datasets show opposite trends. The comparison with the two observational datasets shows that ERA5 generally captures well the temporal and spatial characteristics of the near-surface mean temperature in the Arctic, which makes it suitable to use as a reference in the model evaluations.

Three assessment indices (R, RMSE, and RSTD) were used to evaluate the ability of each CMIP6 model to simulate the spatial pattern of near-surface temperature in the Arctic (Fig. 3). The spatial R is above 0.9 for all models except one. These high correlations indicate that most CMIP6 models perform relatively well in capturing the overall climatological pattern of the Arctic near-surface mean temperature. The RSTD values are close to 1.0 and the RMSE values are less than 0.5 for the spatial variability for all models except one. In general, the CMIP6 models perform well in simulating the spatial variability of near-surface mean temperature in the Arctic, which is also reflected in a small model spread in these metrics for the spatial variability.

The ability of CMIP6 models to simulate the temporal near-surface temperature variations in the Arctic was evaluated by comparing the mean temperature, temperature trend, and temperature interannual standard deviation between the models and the benchmark. According to the differences in climatological near-surface mean temperature during 1979–2014 between the CMIP6 models and ERA5 reanalysis dataset (Fig. 4a), most CMIP6 models underestimate the near-surface mean temperature in the Arctic, and the cold bias of the multimodel ensemble mean (MMEM) of the 22 CMIP6 models is 0.77°C. Spatially, the cold biases in the near-surface temperature simulations are mainly found over the oceans, and some warm biases appear over land (Fig. 5). The cold biases of the CMIP6 models are largest over the Greenland Sea, the Norwegian Sea, the Barents Sea, and the Kara Sea. On a seasonal basis, the spatial patterns of bias are most obvious in winter, and the largest cold (warm) bias occurs near the Greenland Sea (northern Russia), exceeding 5°C (4°C) (not shown). Although the MMEM of the annual temperature trend is almost identical to the trend in ERA5, large differences in temperature trends between the models exist, suggesting large uncertainty in modeling amplified Arctic warming even during the contemporary period. For example, the warming trends are significantly overestimated in EC-Earth3 and NESM3 while significantly underestimated in CAMS-CSM1-0

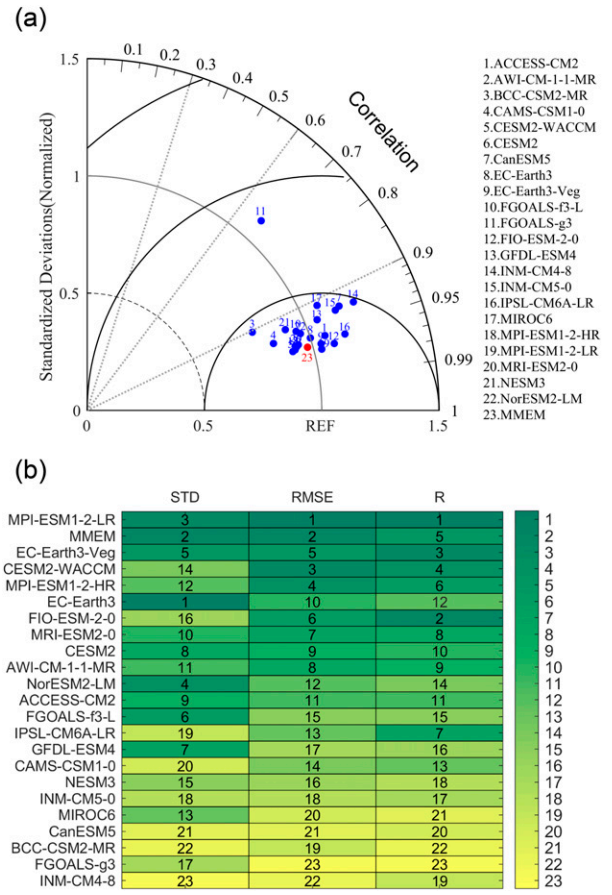

FIG. 3. (a) Taylor diagram of the correlation coefficients (R; azimuthal angle), ratios of standard deviations (RSTD; radial distance from the origin) and normalized centered root-mean-square errors [RMSE; radial distance from reference (REF) on the x axis] between the reference (REF) and simulated near-surface mean temperature in the Arctic during 1979–2014, with ERA5 as the reference. The root-mean-square differences were normalized by dividing by the standard deviations of the reference. Models (represented by points) closer to REF show better agreement with the reference (i.e., R and RSTD close to 1 and RMSE close to 0). The multimodel ensemble mean (MMEM) refers to the average of the 22 CMIP6 models used in this study. (b) Portrait diagram showing the rankings (1–23 with 1 being the top performing) of the models according to R, RMSE, and RSTD.

and FGOALS-g3 (Fig. 4b). In terms of seasons, the intermodel spread is largest in winter and smallest in summer, and the MMEM overestimates the temperature trend in winter and autumn but underestimates the trend in spring (Table 3). According to the results of interannual variability skill, BCC-CSM2-MR and CanESM5 perform best in simulating interannual variations of near-surface temperature in the Arctic, while FGOALS-g3 and EC-Earth3 (whose temperature trends are severely overestimated) perform the worst (Fig. 4c).

Based on the spatial TS (Fig. 6a), the top five models are MPI-ESM1-2-LR, EC-Earth3-Veg, FIO-ESM-2-0, MPI-ESM1-2-HR, and CESM2-WACCM. In terms of temporal TR

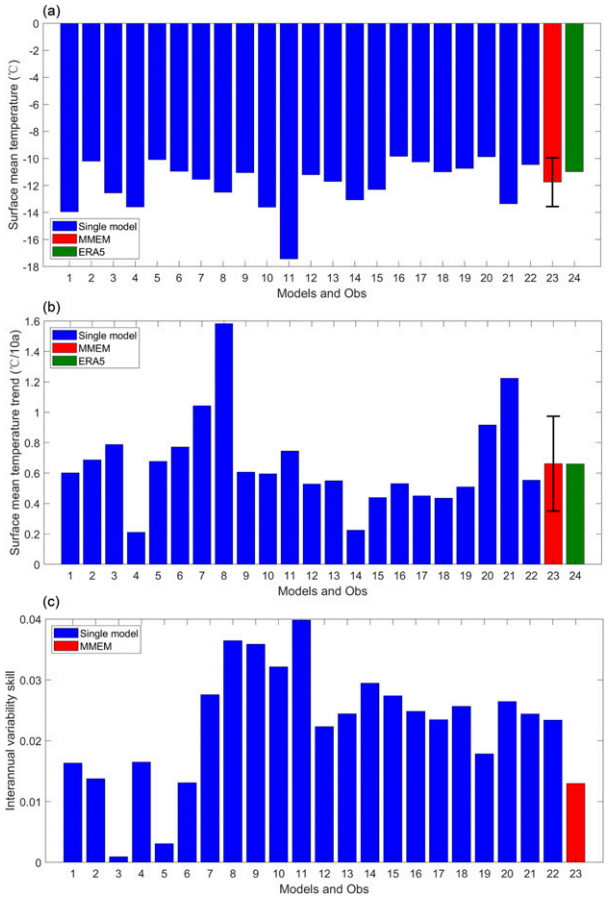

FIG. 4. (a) Near-surface mean temperature (°C), (b) near-surface mean temperature trend (°C decade^{-1}), and (c) interannual variability skill score (IVS) of near-surface mean temperature in the Arctic during 1979–2014 from ERA5, each CMIP6 model and the multimodel ensemble mean (MMEM) of the 22 CMIP6 models. The numbers 1–22 refer to the model names in Table 3. Number 23 is the MMEM, and number 24 is the ERA5 reanalysis dataset. The error bars show one standard deviation of the multimodel ensemble. All trends in (b) are statistically significant at the 0.05 significance level. A IVS closer to 0 in (c) indicates better model performance in terms of interannual variations.

(Fig. 6b), the top five models are CESM2, AWI-CM-1-1MR, NorESM2-LM, FIO-ESM2-0, and MPI-ESM1-2-LR. The relationship between temporal and spatial ranks for each model is shown in Fig. 6c, suggesting the overall best performing models are CESM2, AWI-CM-1-1MR, FIO-ESM-2-0, EC-Earth3-Veg, MPI-ESM1-2-LR, and CESM2-WACCM. There is a relationship between how well the models capture the spatial characteristics and temporal characteristics (R = 0.49), but there are also some models that have a high temporal and low spatial M_R value and vice versa. The models perform better over Arctic land than the ocean (not shown). The MMEM performs better than any single model except for one in terms of spatial characteristics and the MMEM also performs well in terms of temporal characteristics relative to the other models. Thus, the MMEM stands out as a good representation of Arctic near-surface temperature characteristics based on the overall model rankings (Fig. 6c). In the following analysis, we use the MMEM results to assess the projection of Arctic near-surface temperatures in the near-term, midterm, and long-term future.

b. *Projected future temperature from the CMIP6 models*

As shown in Fig. 7, the future Arctic near-surface temperature is projected to increase under all the three scenarios although it begins to slightly decrease by the end of the twenty-first century under SSP1-2.6 scenario. The Arctic's warming rate from 1986 to 2100 is much higher than that of the Northern Hemisphere and the global mean under the three different scenarios (You et al. 2021).

Figure 8 shows the spatial patterns of annual mean near-surface temperature change in the Arctic according to the MMEM for the three periods relative to 1986–2005 under the three scenarios. Projections for the regionally averaged mean near-surface temperature increases in the Arctic under SSP1-2.6, SSP2-4.5, and SSP5-8.5 scenarios are +2.5°, +2.6°, and +2.8°C respectively in the near term (2021–40), +3.3°, +4.0°, and +5.1°C in the midterm (2014–60), and +3.5°, +5.8°, and +10.4°C in the long-term (2081–2100) relative to the reference period based on the CMIP6 MMEM. The spatial pattern is similar across all scenarios, which suggests that Arctic warming has a similar mechanism under different forcings. The strongest warming is found in the Chukchi and Beaufort Seas on the Pacific side and the Barents and Kara Seas on the Atlantic side, which is also reflected in the results of Boeke and Taylor (2018) using CMIP5 models. The warming is the smallest in the Greenland and Norwegian Seas. In terms of temperature trends (Fig. 9), the warming rates of Arctic near-surface temperature in the near term (2021–40) are 0.46°C decade^{-1} under SSP1-2.6, 0.68°C decade^{-1} under SSP2-4.5, and 0.83°C decade^{-1} under SSP5-8.5. Under the SSP1-2.6 scenario, the warming rate gradually decreases with time and becomes negative at the end of the twenty-first century. For the high-emissions SSP5-8.5 scenario, the warming rate continues to increase with time. In the long-term future (2081–2100), the warming rates of Arctic near-surface temperature are −0.11°, 0.37°, and 1.36°C decade^{-1} under SSP1-2.6, SSP2-4.5, and SSP5–8.5, respectively. As shown in Table 4, the warming trend over the central Arctic Ocean (0.95°, 1.13°, and 1.48°C decade^{-1} under SSP1-2.6, SSP2-4.5, and SSP5-8.5, respectively) is higher than that over the surrounding land area (0.66°, 0.94°, and 1.22°C decade^{-1} under SSP1-2.6, SSP2-4.5, and SSP5-8.5, respectively). In the northern Atlantic, the models show a cooling trend in the near-term future for all scenarios, reflecting a phenomenon known as the North Atlantic warming hole (Drijfhout et al. 2012).

c. *Uncertainties in simulated Arctic temperatures in the CMIP6 models*

Internal variability, model uncertainty, and scenario uncertainty are considered sources of uncertainty in climate projections (Hawkins and Sutton 2009, 2011). Figure 10 shows the fractional uncertainty and the fraction of total variance in decadal near-surface mean temperature in the Arctic. Throughout the century, model uncertainty and scenario

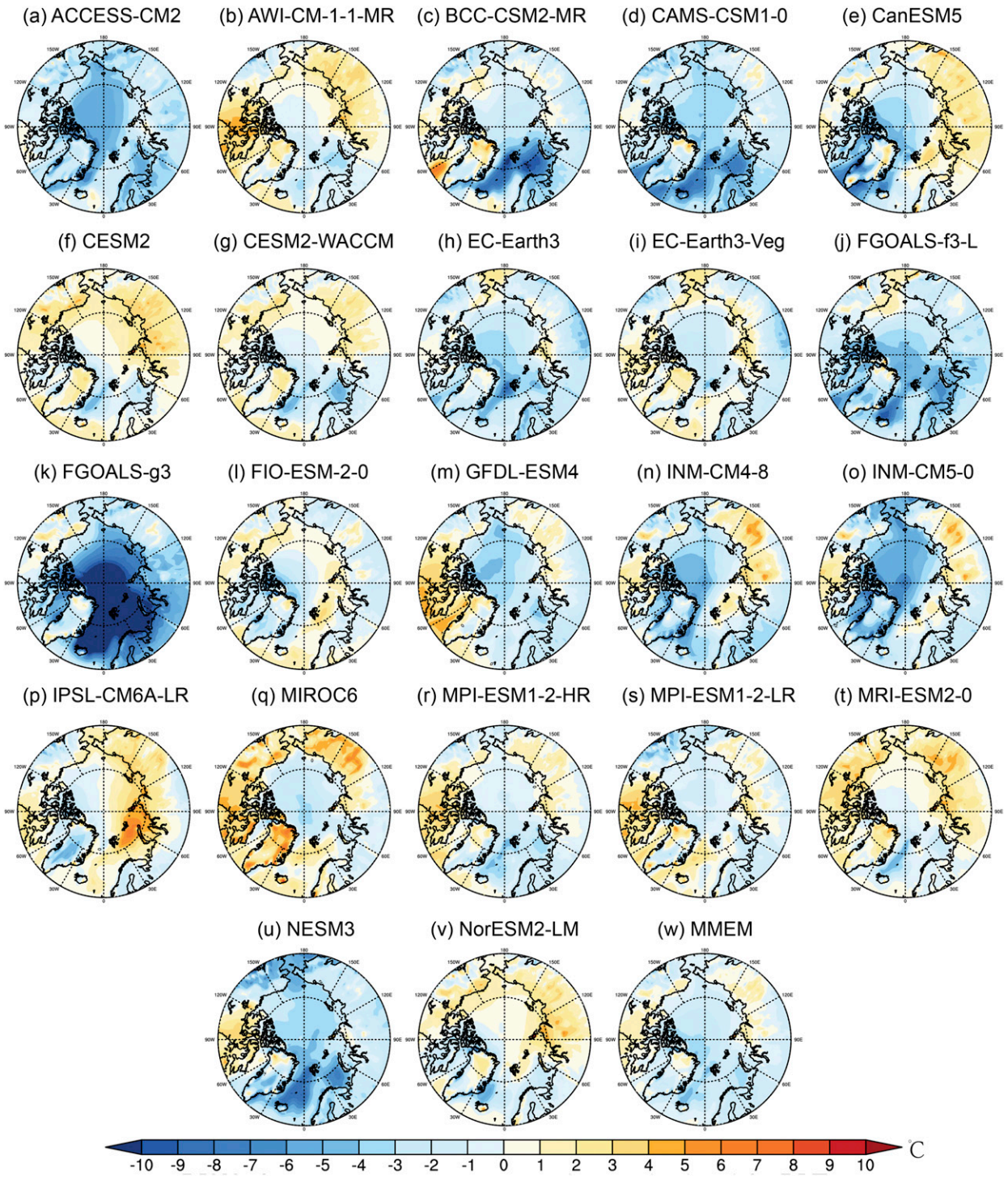

FIG. 5. The difference of climatological near-surface mean temperature during 1979–2014 between the CMIP6 models and the ERA5 reanalysis dataset for (a)–(v) individual CMIP6 models and (w) the multimodel ensemble means (MMEM) of the 22 CMIP6 models.

uncertainty are the major sources of uncertainties for CMIP6 model projections. Model uncertainty is the dominant contributor before the midterm (2014–60) while scenario uncertainty becomes more important at the end of the twenty-first century. The smallest total fractional uncertainty occurs around 2040 associated with the changing dominance of the contributions from model to scenario uncertainty, which is consistent with previous global and regional studies (Hawkins

TABLE 3. Spatially averaged annual and seasonal near-surface temperature mean (°C) and trend (°C decade^{-1}) from the individual CMIP6 models, the multimodel ensemble mean (MMEM) of 22 CMIP6 models, as well as for ERA5 in the Arctic during 1979–2014. Statistically significant trends (95% level) are shown in boldface.

	Trend					Mean				
Model	MAM	JJA	SON	DJF	Annual	MAM	JJA	SON	DJF	Annual
ACCESS-CM2	**0.40**	**0.29**	**0.98**	**0.76**	**0.60**	−16.35	1.68	−12.07	−28.74	−13.94
AWI-CM-1-1-MR	**0.66**	**0.33**	**0.86**	**0.90**	**0.69**	−13.51	3.18	−7.81	−22.34	−10.20
BCC-CSM2-MR	**0.64**	**0.42**	**1.10**	**1.02**	**0.79**	−13.73	1.04	−11.02	−25.99	−12.55
CAMS-CSM1-0	0.32	**0.03**	0.19	**0.51**	**0.21**	−14.20	1.51	−12.93	−29.24	−13.58
CanESM5	**0.89**	**0.50**	**1.42**	**1.33**	**0.68**	−13.44	3.14	−10.30	−24.75	−10.09
CESM2	**0.44**	**0.41**	**1.09**	**0.81**	**0.77**	−12.54	2.08	−7.14	−22.11	−10.95
CESM2-WACCM	**0.51**	**0.39**	**1.14**	**1.11**	**1.04**	−13.73	1.87	−8.04	−23.35	−11.54
EC-Earth3	**1.27**	**0.97**	**2.08**	**2.01**	**1.58**	−14.94	2.22	−11.26	−24.60	−12.49
EC-Earth3-Veg	**0.41**	**0.39**	**0.98**	**0.63**	**0.61**	−14.06	3.00	−9.83	−22.98	−11.05
FGOALS-f3-L	**0.52**	**0.21**	**0.68**	**0.96**	**0.59**	−17.14	2.30	−11.29	−28.49	−13.60
FGOALS-g3	**0.80**	**0.49**	**0.80**	**0.87**	**0.75**	−20.56	0.69	−16.62	−32.77	−17.43
FIO-ESM-2-0	0.02	**0.03**	**0.09**	**0.07**	**0.53**	−14.80	3.52	−8.68	−24.76	−11.20
GFDL-ESM4	**0.37**	**0.29**	**0.83**	**0.79**	**0.55**	−14.14	3.00	−10.08	−26.01	−11.71
INM-CM4-8	0.19	**0.20**	**0.40**	0.08	**0.22**	−14.56	2.33	−13.05	−27.61	−13.07
INM-CM5-0	0.27	**0.33**	**0.67**	**0.67**	**0.44**	−13.63	2.51	−11.55	−25.71	−12.30
IPSL-CM6A-LR	**0.37**	**0.31**	**0.86**	**0.59**	**0.53**	−13.01	4.23	−6.92	−23.37	−9.84
MIROC6	**0.38**	**0.19**	**0.63**	**0.63**	**0.45**	−14.05	5.05	−8.69	−22.54	−10.25
MPI-ESM1-2-HR	**0.45**	**0.26**	**0.60**	0.44	**0.44**	−13.69	2.90	−8.57	−23.55	−11.00
MPI-ESM1-2-LR	**0.54**	**0.23**	**0.67**	**0.53**	**0.51**	−13.62	2.90	−8.75	−23.11	−10.74
MRI-ESM2-0	**0.79**	**0.39**	**1.23**	**1.20**	**0.92**	−12.58	3.55	−7.39	−22.18	−9.88
NESM3	**1.35**	**0.46**	**1.31**	**1.76**	**1.22**	−17.99	2.37	−9.35	−27.37	−13.35
NorESM2-LM	**0.65**	**0.30**	**0.76**	**0.73**	**0.55**	−12.68	1.93	−7.96	−22.54	−10.46
MMEM	**0.56 ± 0.31**	**0.35 ± 0.19**	**0.91 ± 0.42**	**0.85 ± 0.45**	**0.66 ± 0.32**	−14.46 ± 1.90	2.61 ± 1.00	−9.93 ± 2.30	−25.16 ± 2.80	−11.75 ± 1.80
ERA5	**0.68**	**0.29**	**0.92**	**0.71**	**0.66**	−13.25	3.43	−9.41	−22.68	−10.98

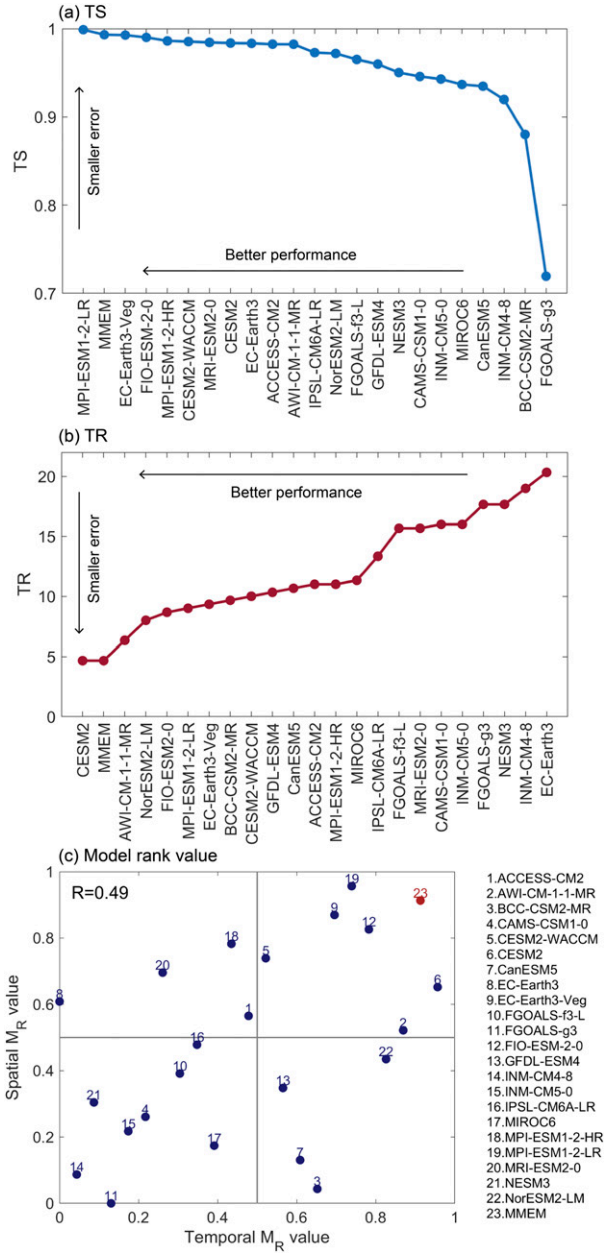

FIG. 6. (a) Taylor skill scores (TS) and (b) temporal rank scores (TR) of near-surface mean temperature in the Arctic. A TS closer to 1 indicates better model performance in terms of spatial variability and a TR closer to 0 indicates better model performance in terms of temporal variations. MMEM is the multimodel ensemble mean. (c) Scatter diagram showing the correlation between temporal and spatial model rank (M_R) values, with higher M_R values indicating higher skill. Each dot represents a model with the same corresponding number displayed to the right of the diagram. The correlation coefficient between temporal and spatial M_R values is 0.49.

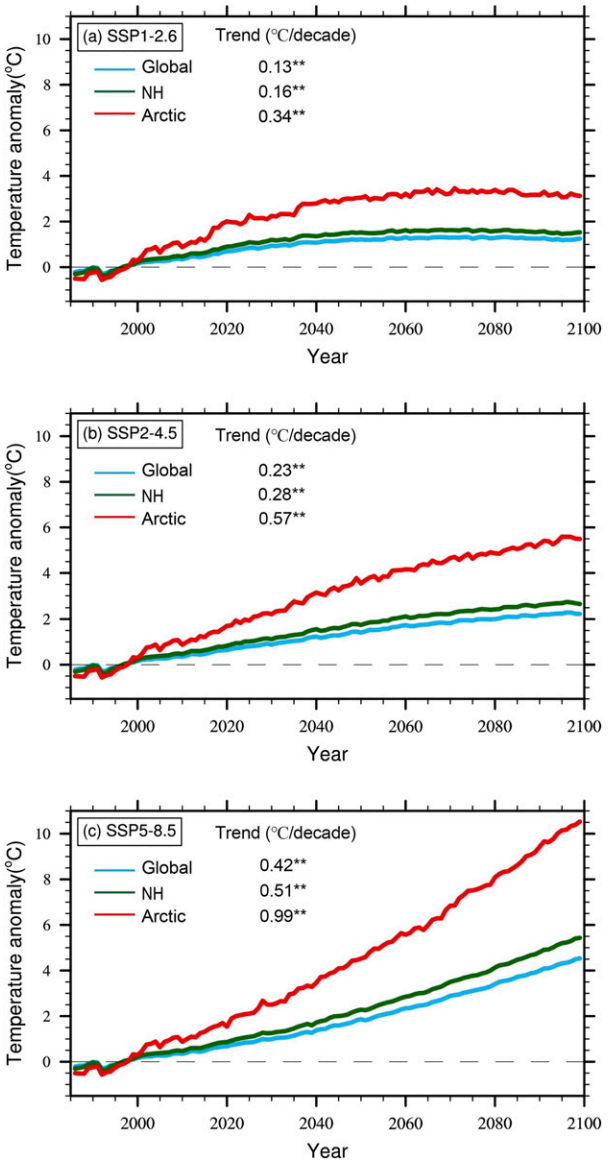

FIG. 7. Time series of Arctic, Northern Hemisphere, and global near-surface mean temperature during 1986–2100 relative to the period 1986–2005 from the multimodel ensemble mean (MMEM) of the 22 CMIP6 models under the three Shared Socioeconomic Pathways (SSPs), (a) SSP1-2.6, (b) SSP2-4.5, and (c) SSP5-8.5.

and Sutton 2009, 2011; Hodson et al. 2013; Wu et al. 2020). Compared with other regions and global mean results, the proportion of model uncertainty to total uncertainty is greater in the Arctic and continues to play an important role at the end of the twenty-first century in spite of the growing scenario uncertainty (Hawkins and Sutton 2009).

The contribution to total uncertainty of both internal variability and model uncertainty decrease while that of scenario uncertainty increases over time. This can be explained by the increasing importance of the anthropogenic forcing and definition of the fractional uncertainty. The time series of the numerator and denominator in the fractional uncertainty formula for each source of uncertainty is shown in Fig. 11. It is clear that all numerators and denominators increase with time, except the numerator of the fractional uncertainty for internal

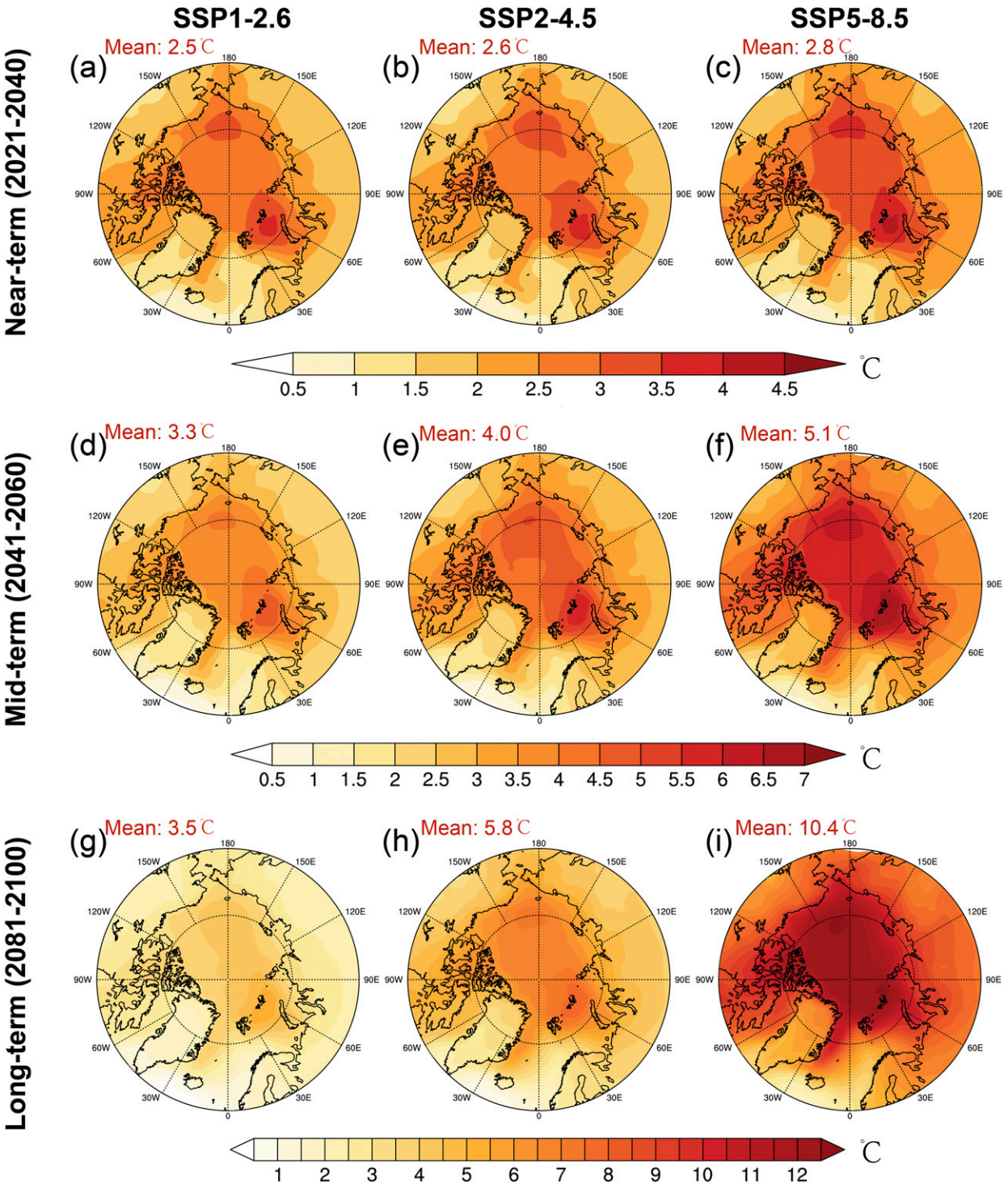

FIG. 8. Spatial patterns of near-surface mean temperature change in the Arctic in the (a)–(c) near-term (2021–40), (d)–(f) midterm (2041–60), and (g)–(i) long-term (2081–2100) relative to the reference period (1986–2005). The projections were obtained from the multimodel ensemble mean of the 22 CMIP6 models under the (left) SSP1-2.6, (center) SSP2-4.5, and (right) SSP5-8.5 scenarios.

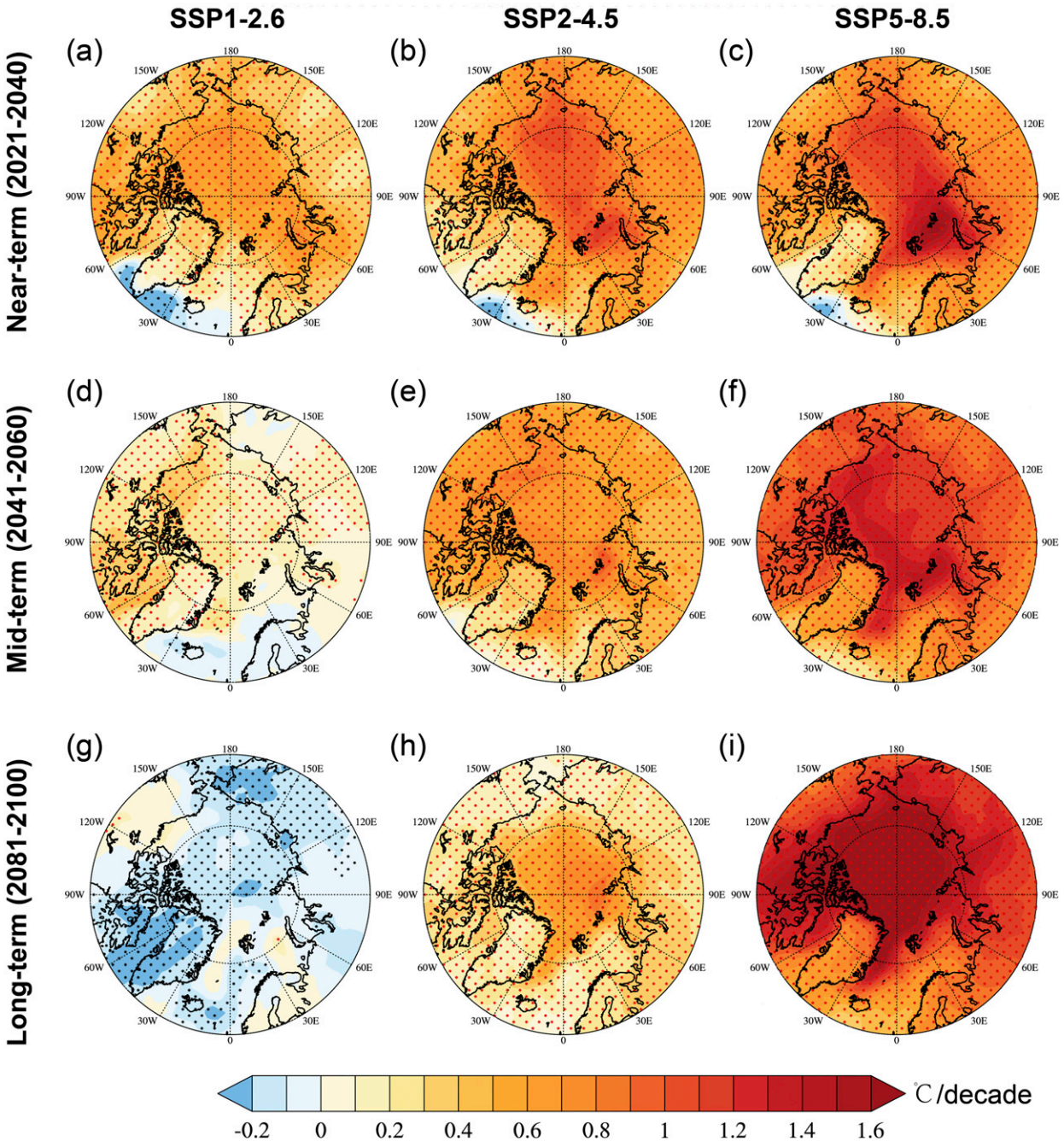

FIG. 9. Spatial patterns of near-surface mean temperature trend in the Arctic in the (a)–(c) near-term (2021–40), (d)–(f) midterm (2041–60), and (g)–(i) long-term (2081–2100). The projections were obtained from the multimodel ensemble mean of the 22 CMIP6 models under the (left) SSP1-2.6, (center) SSP2-4.5, and (right) SSP5-8.5 scenarios.

variability, which is nearly constant by definition. In the case of fractional uncertainty in internal variability and model uncertainty, the denominator becomes increasingly larger than the numerator over time, resulting in decreasing fractional uncertainties with time. For the fractional scenario uncertainty, the distance between the larger denominator and smaller numerator initially grows with time; however, after around 2055 the numerator increases faster than the denominator, thus leading to an increase in the growth rate of the fractional scenario uncertainty with time after 2055. The total fractional uncertainty shows a similar behavior as the fractional scenario uncertainty except that the scenario uncertainty accelerates relative to total uncertainty at the end of the century. Unlike traditional methods, the fractional uncertainty for

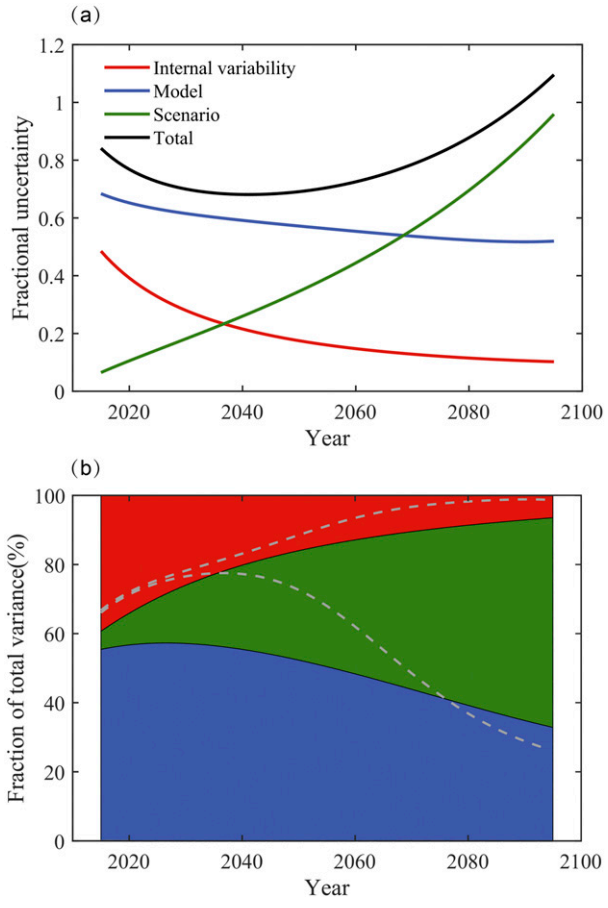

FIG. 10. (a) Fractional uncertainty (the 90% confidence level divided by the mean prediction) and (b) fraction of total variance in decadal near-surface mean predictions in the Arctic. The dotted lines are the result of estimation using single model initial-condition large ensembles (SMILEs) from six models.

internal variability is not constant in time when we use SMILEs for estimation. Figure 10b shows the difference between the conclusions of the two methods. The proportion of internal variability estimated from the models to the total uncertainty (33.24% in 2015) is smaller than the results of the traditional method (39.31% in 2015), and the influence of the internal variability has basically disappeared at the end of the twenty-first century. Since SMILEs are still not widespread in CMIP6 and the ensemble numbers for most models used are too small at this time, more research is needed to make a reliable model uncertainty estimate in the future.

We further examined the spatial distributions of sources of uncertainty in the different periods, which are shown in Fig. 12. Model uncertainty is the dominant source of uncertainty in the Arctic in the near-term and midterm periods and is of comparable importance to scenario uncertainty even at the end of the century in some regions. In general, the uncertainty on the Atlantic side is greater than that on the Pacific side. The Greenland Sea, Norwegian Sea, Kara Sea, and Barents Sea have the largest total fractional uncertainty, and the model uncertainty in these areas dominate the total uncertainty throughout the century. Figure 13 shows the spatial patterns of signal-to-noise ratios for each scenario and period. The smallest signal-to-noise ratio is found in the Greenland Sea, the Norwegian Sea, and the Barents Sea where the values are less than 1.0, indicating that the reliability of the model projection results are very low and large inconsistencies exist between the models, which is consistent with the largest fractional uncertainty in these regions. Elsewhere in the Arctic, the signal-to-noise ratios are consistently larger than 1.0, especially over the Arctic land areas where the signal-to-noise ratio is greater than 3.0, indicating that the models have high confidence in projecting the Arctic land temperature.

4. Discussion

a. *Probable causes for the cold biases in the CMIP6 models*

The ability of different GCMs to simulate the near-surface mean temperature in the Arctic region varies greatly due to different model structures and diverse parameterization schemes (Knutti 2008). The MMEM of the 22 CMIP6 models can mask the deficiency of any single model simulation. Simulated near-surface mean temperature from the MMEM during 1979–2014 is 0.77°C colder than the observationally

TABLE 4. Projected warming over the Arctic Ocean and Arctic land from the multimodel ensemble mean (MMEM) of 22 CMIP6 models in the near-term (2021–40), midterm (2041–60), and long-term (2081–2100) periods and the temperature change relative to 1986–2005 from the MMEM of 22 CMIP6 models under SSP1-2.6, SSP2-4.5, and SSP5-8.5, respectively. Statistically significant trends (95% level) are shown in boldface.

	SSP1-2.6		SSP2-4.5		SSP5-8.5	
	Ocean	Land	Ocean	Land	Ocean	Land
Change (°C)						
Near-term	2.53	2.02	2.61	2.05	2.89	2.31
Mid-term	3.36	2.59	4.03	3.15	5.11	4.03
Long-term	3.62	2.68	5.86	4.55	10.22	8.33
Trend (°C decade^{-1})						
Near-term	**0.50**	**0.40**	**0.78**	**0.56**	**0.96**	**0.66**
Mid-term	**0.18**	**0.16**	**0.62**	**0.52**	**1.13**	**0.94**
Long-term	−0.10	−0.12	**0.44**	**0.28**	**1.48**	**1.22**

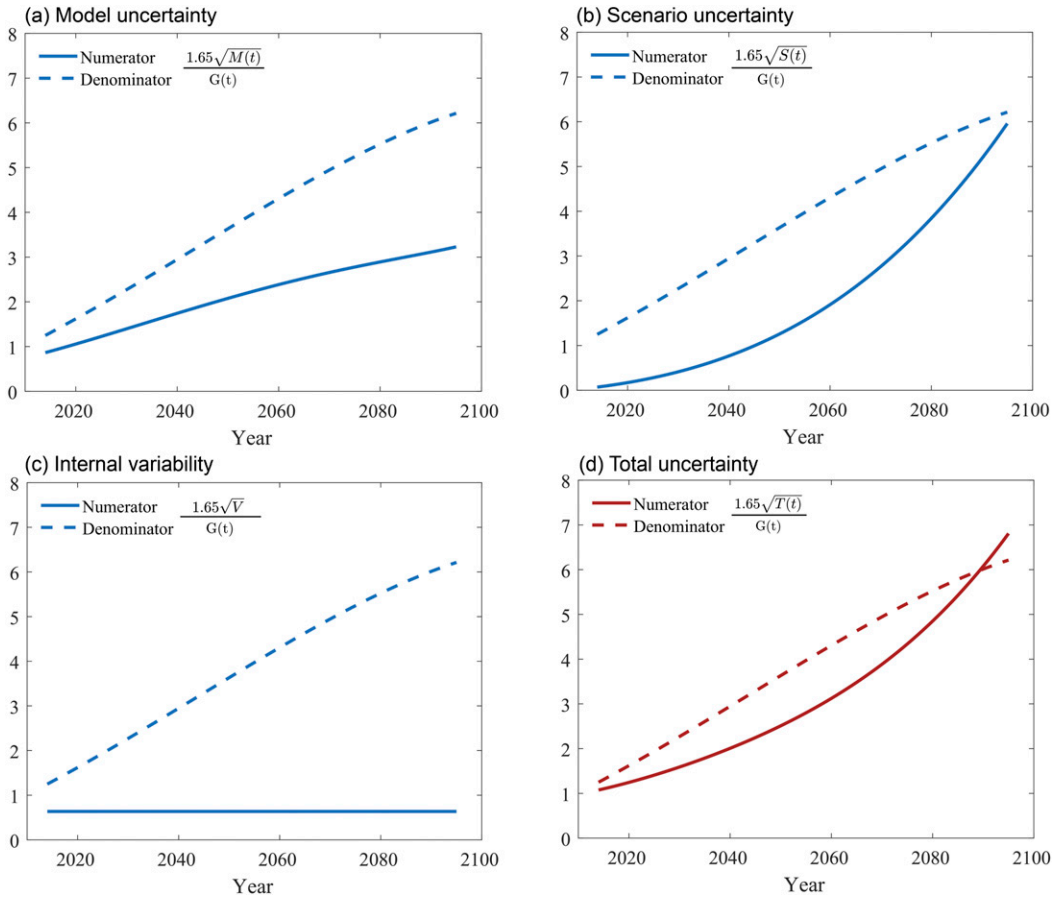

FIG. 11. Time series of the numerator (solid lines) and denominator (dashed lines) of the fractional uncertainty for (a) model uncertainty, (b) scenario uncertainty, (c) internal variability, and (d) the total uncertainty. Here, $M(t)$ is the multiscenario mean model uncertainty, $S(t)$ is the multimodel mean scenario uncertainty, V is the multimodel mean of the variances of the residuals, and $G(t)$ is the mean change of all the predictions.

constrained ERA5 reanalysis with a cold bias maximum of 3°–4°C in the Norwegian Sea, the Barents Sea, and the Kara Sea. Such cold bias was also found in Davy and Outten (2020), who showed that both the CMIP6 and CMIP5 models are more than 1 K colder than ERA5 in the annual mean. In terms of regions, cold biases mainly occur over the Arctic Ocean. These findings indicate that the CMIP6 models still struggle with correctly reproducing ocean–atmosphere processes in regions with frequent heat flux exchange, which has previously been reported for CMIP3 and CMIP5 models (Chapman and Walsh 2007; Hao et al. 2018; Liu et al. 2008). One possible reason for the cold bias over the Arctic Ocean is a poor simulation of sea ice. Correlation analysis of the Arctic Ocean temperature bias in the CMIP6 models and the simulated sea ice area in these models shows that there is a significant negative correlation between the temperature simulation bias and sea ice area. Compared with the result from the National Snow and Ice Data Center (NSIDC) and the Hadley Centre (HadISST1 ice), most models overestimate the area of sea ice (Fig. 14). Previous studies have also found an overestimation of Arctic sea ice extent and thickness in both CMIP5 and CMIP6 models, especially during the annual maximum in March (Davy and Outten 2020; Shu et al. 2020; Wu et al. 2019). Sea ice can affect near-surface temperature through two local feedbacks. One is the high albedo effect, and it affects the atmospheric temperature through the albedo feedback mechanism during the warm seasons. The other is the insulating properties of ice ice, which can isolate the heat exchange between the ocean and the atmosphere, which is more dominant during the cold season (Boeke and Taylor 2018). Therefore, a relatively extensive sea ice cover in the models can cause more shortwave radiation to be reflected back into space during the warm season and a thicker sea ice cover reduces the heat released by the Arctic Ocean into the atmosphere during the cold season, leading to an underestimation of Arctic near-surface temperature. In addition, poleward moisture transport is a significant contributor to Arctic sea ice and temperature variability during the cold season, especially in the Barents, Kara, and Greenland Seas (Woods and Caballero 2016). This suggests that the accurate simulation of poleward moisture transport cannot be ignored when we consider the reasons for the temperature simulation bias particularly in the cold season.

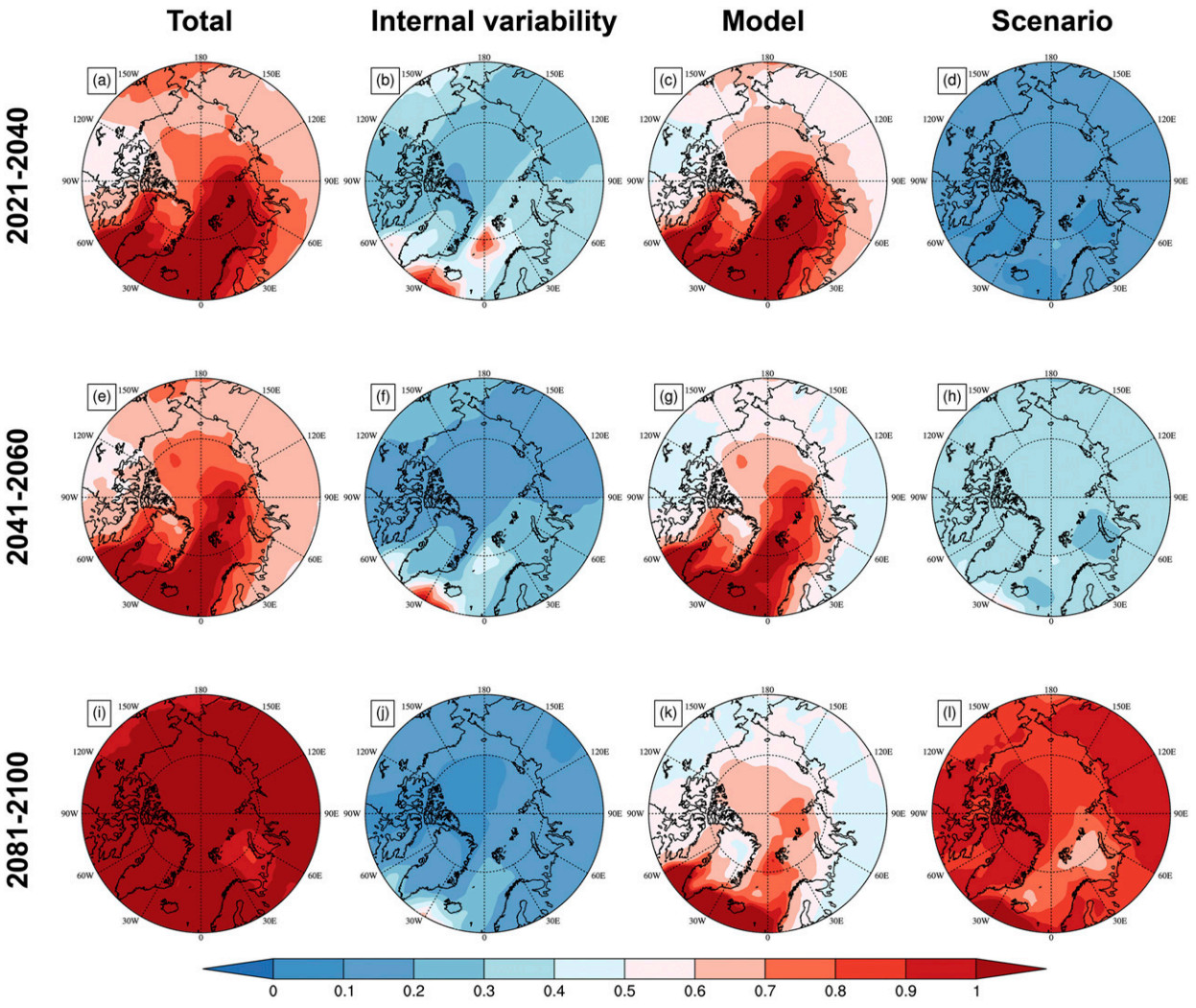

FIG. 12. Spatial patterns of (first column) the total fractional uncertainty, and fractional uncertainty for (second column) internal variability, (third column) model uncertainty, and (fourth column) scenario uncertainty in the Arctic in the (a)–(d) near-term (2021–40), (e)–(h) midterm (2041–60), and (i)–(l) long-term (2081–2100) from the multimodel ensemble mean (MMEM) under the SSP1-2.6, SSP2-4.5, and SSP5-8.5 scenarios.

A previous study using the CMIP3 and CMIP5 models concluded that the models exhibit large uncertainty in the simulation of the Atlantic meridional overturning circulation (AMOC) (Reintges et al. 2017), which may lead to large intermodel differences in the simulation of ocean heat transport in the Northern Atlantic region. One possible reason for the overestimation of sea ice is that the northward ocean heat transport is underestimated. As shown in Fig. 14, models with a weak Atlantic poleward heat transport have more extensive sea ice and greater ocean temperature cold biases. A near-constant influx of relatively warm Atlantic surface water keeps the Norwegian Sea and nearby areas generally ice-free in the present. But the sea ice simulated by the CMIP6 models is generally too high in these areas, leading to the largest temperature biases compared to other model biases (Notz et al. 2020; Levang and Schmitt 2020; Mahajan et al. 2011; Shu et al. 2020).

b. *Improvements and changes in the CMIP6 models*

An encouraging increase in credibility and improvement in the simulation of mean states, interannual variability, and past climate evolution is evident in the progression from the CMIP3/5 to CMIP6 models (Eyring et al. 2016). Improved model performances are due to higher spatial resolution, and advances in simulating certain physical processes (T. Zhou et al. 2019). Table 5 summarizes the comparison between the Arctic surface temperature simulated by models such as those in CMIP3/CMIP5 in the past and the results of observations or reanalysis. Previous models have generally underestimated the near-surface mean temperature of the

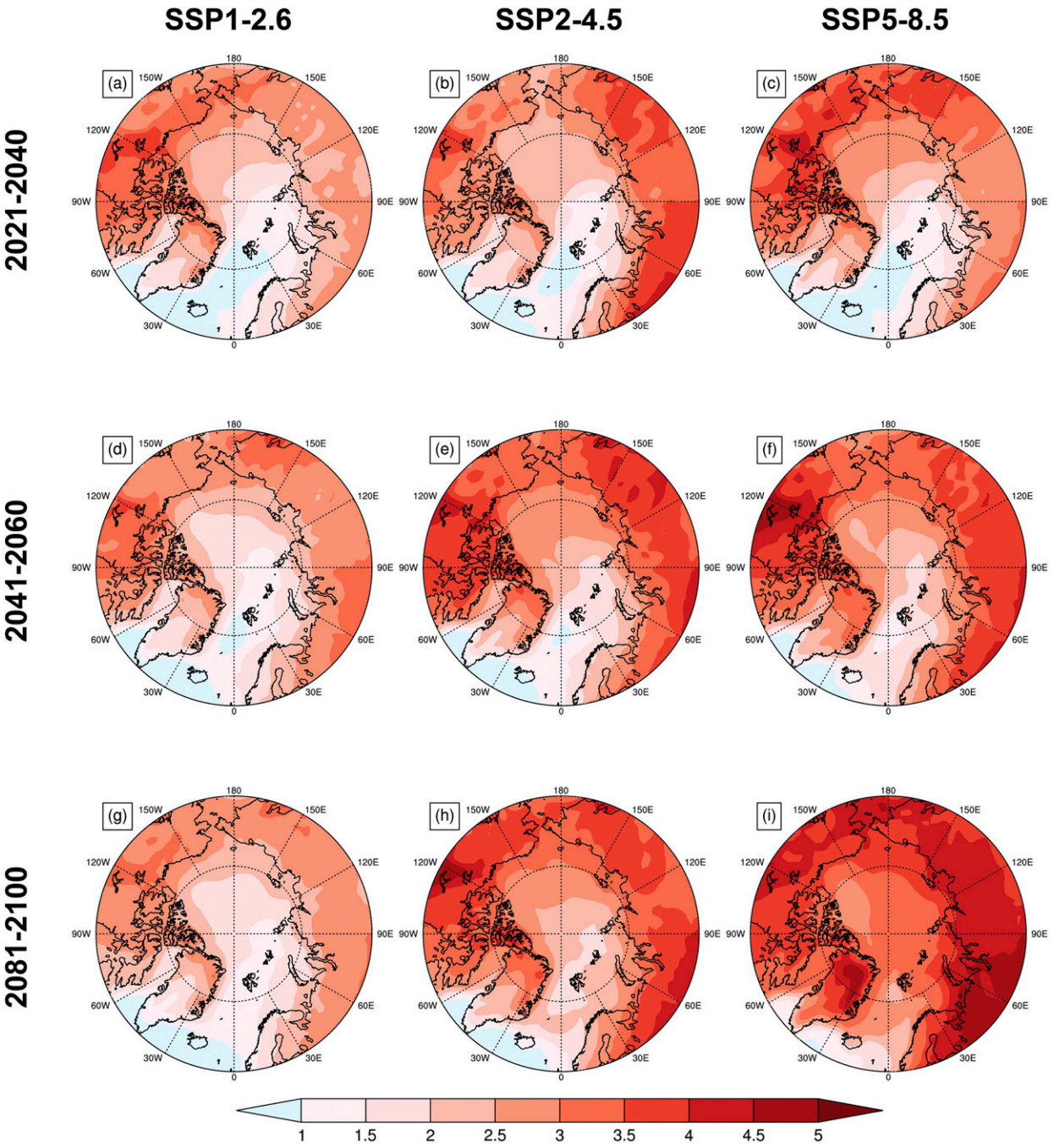

FIG. 13. Spatial patterns of the signal-to-noise ratio in the Arctic in the (a)–(c) near-term (2021–40), (d)–(f) midterm (2041–60), and (g)–(i) long-term (2081–2100) from the multimodel ensemble mean (MMEM) under the (left) SSP1-2.6, (center) SSP2-4.5, and (right) SSP5-8.5 scenarios.

Arctic, and the annual mean temperature in recent decades has been underestimated by 1°–2°C with the larger errors in winter and spring. Comparing the results from previous studies with those of this study, the CMIP6 models have improved the Arctic near-surface temperature biases, which may be related to the better representation of the sea ice extent/thickness and advance/retreat in the CMIP6 models (Davy and Outten 2020).

The CMIP6 models project much stronger warming than CMIP5 at both the global and regional scales (Almazroui et al. 2020; Bracegirdle et al. 2020; Tokarska et al. 2020; Wyser et al. 2020). This may be due to higher equilibrium climate sensitivity

TABLE 5. Summary of studies comparing simulated and observed warming in the Arctic region.

Region	Dataset		Study period	Results		Reference
	Observations	Models		Season	Results	
Arctic	ERA5	49 CMIP5 models 34 CMIP6 models	1979–2004	January and February	Underestimated by more than −4°C	Davy and Outten (2020)
Arctic	GISTEMP/MLOST/ HadCRUT4/GHCN-M	41 CMIP5 models	1955–2004	Cold season	Models overestimated the trend for increasing SAT	Xie et al. (2016)
Arctic	HadCRUT4.6	36 CMIP5 models	1880–2017	Annual	Overestimated the anthropogenically induced secular warming rate	Huang et al. (2019)
Arctic	ERA-Interim	32 CMIP5 models	1900–2005	Annual	Cold bias: −0.04°C (SAT)	Hao et al. (2018)
Arctic	ERA-40	14 CMIP3 models	1981–2000	Annual	Cold bias: from −2° to −1°C (SAT)	Chapman and Walsh (2007)
Northern Eurasia/ northern North America	NOAA	19 AMIP AGCMs	1979–88	Spring	Cold bias: from −3.3° to −2°C (SST)	Tao et al. (1996)
Arctic	ERA-Interim/GSOD (210 stations)	HIRHAM5 regional climate model	1979–2014	Spring Summer Autumn Winter Annual	Cold bias: −2.18°C (SAT) Cold bias: −1.82°C (SAT) Cold bias: −1.93°C (SAT) Cold bias: −1.85°C (SAT) Cold bias: −1.95°C (SAT)	X. Zhou et al. (2019)
Barents Sea–Kara Sea	ERA-Interim	25 CMIP5 models	1979–2005	Winter	Cold bias: −8°C (SAT)	Wang et al. (2020)
Arctic	ERA-Interim	Global coupled climate model EC-Earth2.3	1980–99	Annual	Cold bias: −2°C (SAT)	Koenigk et al. (2013)
Arctic	ERA-Interim	Global coupled climate model CAMS-CSM	1979–2013	Annual	Underestimate mean temperature and overestimate the linear trend (SAT)	Wei et al. (2018)

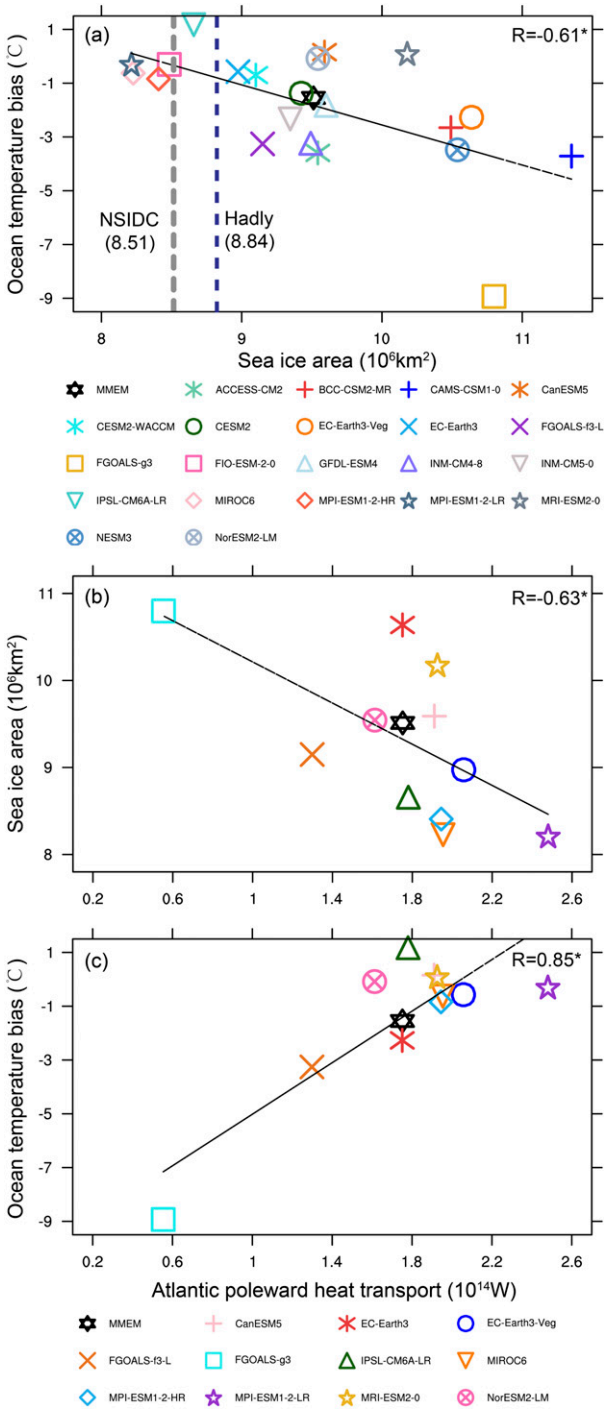

FIG. 14. (a) Correlation between the annual mean ocean surface temperature bias (CMIP6 minus ERA5) in the Arctic for 1979–2014 and the simulated annual mean sea ice area for the same period for all CMIP6 models and the multimodel ensemble mean (MMEM) of all models. The gray dotted line represents the observed value from the National Snow and Ice Data Center, downloaded from https://nsidc.org/data/G02202/versions/3. The blue dotted line represents the observed value from the Hadley Center, downloaded from https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html. (b),(c) Correlation between the annual mean sea ice area/annual mean ocean surface temperature bias (CMIP6 minus ERA5) in the Arctic for 1979–2014 and the simulated mean Atlantic poleward heat transport for the same period for all CMIP6 models and the MMEM. Sea ice area was computed from SIC (defined as the area with SIC > 15% in the north of 66°N). The poleward heat transport is defined as the poleward flux at 70°N. In all panels, *R* denotes the correlation coefficient, and an asterisk (*) indicates that the correlation is significant at the 0.05 significance level.

values in multiple GCMs submitted to CMIP6 (Zelinka et al. 2020). In fact, approximately one-third of the submitted CMIP6 models have estimated equilibrium climate sensitivity ranges that exceed the 1.5°–4.5°C range (17%–83% range) reported by IPCC AR5 (Tokarska et al. 2020). Therefore, it is very important to investigate whether the future warming trends projected by these models are realistic.

5. Conclusions

In this study, the simulated Arctic near-surface temperatures from the 22 CMIP6 models are evaluated against the observationally constrained ERA5 reanalysis dataset. Additionally, we assessed the near-surface mean temperatures and temperature trends in the Arctic in the near-term, midterm, and long-term future under three SSPs. The uncertainty in the temperature projections was quantified and broken down into three components: uncertainty due to internal variability of the climate system, model uncertainty, and scenario uncertainty due to large spreads in future emissions and socioeconomic pathways.

During 1979–2014, the 22 CMIP6 models that we examined underestimate the temperature in the Arctic with cold biases relative to the ERA5 reanalysis dataset, which are especially evident over the Greenland Sea, the Barents Sea, and the Kara Sea. The cold bias is mainly focused over the Arctic Ocean, which is most obvious during the cold season, and the land area shows a general warm bias. The models generally do a good job in representing the spatial variability of climatological mean near-surface temperature in the Arctic, while there are some differences between models in the simulation of temporal variations. In general, models that accurately simulate spatial variability also perform well in simulating temporal variability.

The CMIP6 models project that the Arctic warming will continue in the future under the SSP1-2.6, SSP2-4.5, and SSP5-8.5 scenarios and that the warming rate in the Arctic will be expected to be more than double the Northern Hemisphere and global means. Given that the CMIP6 models simulate a much stronger Arctic warming than the CMIP5 models, future studies should assess how realistic these warming trends are, for example through the use of emergent constraints (Hall et al. 2019).

Our results show that there are large uncertainties in the projections of Arctic near-surface mean temperature changes from the 22 CMIP6 models under the three SSPs. In general, model uncertainty and scenario uncertainty are the major sources of uncertainty, accounting for 52.3% and 31.7% of the total uncertainty respectively in 2050, and 32.9% and 60.7% of

the total uncertainty respectively by the end of the twenty-first century. Model uncertainty is always the dominant uncertainty component in oceanic regions that are affected by heat transport from the AMOC. More detailed investigations of model uncertainty and warming mechanisms in the Arctic region are needed to better understand the sources of model uncertainty and to reduce the overall uncertainty in the projections, which is also the direction of future research.

Acknowledgments. This study is supported by the National Key R&D Program of China (2017YFA0603804) and the National Natural Science Foundation of China (41971072, 41771069, 91937302, and 41911530187). J. Cohen is supported by the U.S. National Science Foundation Grant PLR-1901352, and D. Chen is supported by the Swedish BECC and MERGE. We are very grateful to the three reviewers and Dave Bonan for their constructive comments and thoughtful suggestions.

REFERENCES

Almazroui, M., S. Saeed, F. Saeed, M. N. Islam, and M. Ismail, 2020: Projections of precipitation and temperature over the South Asian countries in CMIP6. *Earth Syst. Environ.*, **4**, 297–320, https://doi.org/10.1007/s41748-020-00157-7.

Barton, N. P., S. A. Klein, and J. S. Boyle, 2014: On the contribution of longwave radiation to global climate model biases in Arctic lower tropospheric stability. *J. Climate*, **27**, 7250–7269, https://doi.org/10.1175/JCLI-D-14-00126.1.

Blackport, R., and J. A. Screen, 2020: Weakened evidence for mid-latitude impacts of Arctic warming. *Nat. Climate Change*, **10**, 1065–1066, https://doi.org/10.1038/s41558-020-00954-y.

Boeke, R. C., and P. C. Taylor, 2018: Seasonal energy exchange in sea ice retreat regions contributes to differences in projected Arctic warming. *Nat. Commun.*, **9**, 5017, https://doi.org/10.1038/s41467-018-07061-9.

Bonan, D. B., K. C. Armour, G. H. Roe, N. Siler, and N. Feldl, 2018: Sources of uncertainty in the meridional pattern of climate change. *Geophys. Res. Lett.*, **45**, 9131–9140, https://doi.org/10.1029/2018GL079429.

Box, J. E., and Coauthors, 2019: Key indicators of Arctic climate change: 1971–2017. *Environ. Res. Lett.*, **14**, 045010, https://doi.org/10.1088/1748-9326/aafc1b.

Bracegirdle, T. J., and Coauthors, 2020: Twenty first century changes in Antarctic and Southern Ocean surface climate in CMIP6. *Atmos. Sci. Lett.*, **21**, e984, https://doi.org/10.1002/asl.984.

Chapman, W. L., and J. E. Walsh, 2007: Simulations of Arctic temperature and pressure by global coupled models. *J. Climate*, **20**, 609–632, https://doi.org/10.1175/JCLI4026.1.

Chen, W., Z. Jiang, and L. Li, 2011: Probabilistic projections of climate change over China under the SRES A1B scenario using 28 AOGCMs. *J. Climate*, **24**, 4741–4756, https://doi.org/10.1175/2011JCLI4102.1.

Cohen, J., and Coauthors, 2014: Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.*, **7**, 627–637, https://doi.org/10.1038/ngeo2234.

——, and Coauthors, 2020: Divergent consensuses on Arctic amplification influence on midlatitude severe winter weather. *Nat. Climate Change*, **10**, 20–29, https://doi.org/10.1038/s41558-019-0662-y.

Cowtan, K., and R. G. Way, 2014: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart. J. Roy. Meteor. Soc.*, **140**, 1935–1944, https://doi.org/10.1002/qj.2297.

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.

Deng, K., X. Jiang, C. Hu, and D. Chen, 2020: More frequent summer heat waves in southwestern China linked to the recent declining of Arctic sea ice. *Environ. Res. Lett.*, **15**, 074011, https://doi.org/10.1088/1748-9326/ab8335.

Dodd, E. M. A., C. J. Merchant, N. A. Rayner, and C. P. Morice, 2015: An investigation into the impact of using various techniques to estimate arctic surface air temperature anomalies. *J. Climate*, **28**, 1743–1763, https://doi.org/10.1175/JCLI-D-14-00250.1.

Drijfhout, S., G. J. van Oldenborgh, and A. Cimatoribus, 2012: Is a decline of AMOC causing the warming hole above the North Atlantic in observed and modeled warming patterns? *J. Climate*, **25**, 8373–8379, https://doi.org/10.1175/JCLI-D-12-00490.1.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Gidden, M. J., 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.

Goosse, H., and Coauthors, 2018: Quantifying climate feedbacks in polar regions. *Nat. Commun.*, **9**, 1919, https://doi.org/10.1038/s41467-018-04173-0.

Graham, R. M., and Coauthors, 2019: Evaluation of six atmospheric reanalyses over Arctic sea ice from winter to early summer. *J. Climate*, **32**, 4121–4143, https://doi.org/10.1175/JCLI-D-18-0643.1.

Hall, A., P. Cox, C. Huntingford, and S. Klein, 2019: Progressing emergent constraints on future climate change. *Nat. Climate Change*, **9**, 269–278, https://doi.org/10.1038/s41558-019-0436-6.

Hao, M., J. Huang, Y. Luo, X. Chen, Y. Lin, Z. Zhao, and Y. Xu, 2018: Narrowing the surface temperature range in CMIP5 simulations over the Arctic. *Theor. Appl. Climatol.*, **132**, 1073–1088, https://doi.org/10.1007/s00704-017-2161-2.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1108, https://doi.org/10.1175/2009BAMS2607.1.

——, and ——, 2011: The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dyn.*, **37**, 407–418, https://doi.org/10.1007/s00382-010-0810-6.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hodson, D. L. R., S. P. E. Keeley, A. West, J. Ridley, E. Hawkins, and H. T. Hewitt, 2013: Identifying uncertainties in Arctic climate change projections. *Climate Dyn.*, **40**, 2849–2865, https://doi.org/10.1007/s00382-012-1512-z.

Hu, X., H. Fan, M. Cai, S. A. Sejas, P. Taylor, and S. Yang, 2020: A less cloudy picture of the inter-model spread in future global warming projections. *Nat. Commun.*, **11**, 4472, https://doi.org/10.1038/s41467-020-18227-9.

Huang, J., and Coauthors, 2017: Recently amplified Arctic warming has contributed to a continual global warming trend. *Nat. Climate Change*, **7**, 875–879, https://doi.org/10.1038/s41558-017-0009-5.

——, T. H. Ou, D. L. Chen, Y. Luo, and Z. C. Zhao, 2019: The amplified Arctic warming in the recent decades may have been

overestimated by CMIP5 models. *Geophys. Res. Lett.*, **46**, 13 338–13 345, https://doi.org/10.1029/2019GL084385.

IPCC, 2007: *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 996 pp.

——, 2013: *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, 1535 pp., https://doi.org/10.1017/CBO9781107415324.

Jiang, Z., W. Li, J. Xu, and L. Li, 2015: Extreme precipitation indices over China in CMIP5 models. Part I: Model evaluation. *J. Climate*, **28**, 8603–8619, https://doi.org/10.1175/JCLI-D-15-0099.1.

Karlsson, J., and G. Svensson, 2013: Consequences of poor representation of Arctic sea-ice albedo and cloud-radiation interactions in the CMIP5 model ensemble. *Geophys. Res. Lett.*, **40**, 4374–4379, https://doi.org/10.1002/grl.50768.

Knutti, R., 2008: Should we believe model predictions of future climate change? *Philos. Trans. Roy. Soc.*, **366A**, 4647–4664, https://doi.org/10.1098/rsta.2008.0169.

Koenigk, T., L. Brodeau, R. G. Graversen, J. Karlsson, G. Svensson, M. Tjernström, U. Willén, and K. Wyser, 2013: Arctic climate change in 21st century CMIP5 simulations with EC-Earth. *Climate Dyn.*, **40**, 2719–2743, https://doi.org/10.1007/s00382-012-1505-y.

Landrum, L., and M. M. Holland, 2020: Extremes become routine in an emerging new Arctic. *Nat. Climate Change*, **10**, 1108–1115, https://doi.org/10.1038/s41558-020-0892-z.

Lehner, F., C. Deser, N. Maher, J. Marotzke, E. M. Fischer, L. Brunner, R. Knutti, and E. Hawkins, 2020: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth Syst. Dyn.*, **11**, 491–508, https://doi.org/10.5194/esd-11-491-2020.

Lenssen, N. J. L., G. A. Schmidt, J. E. Hansen, M. J. Menne, A. Persin, R. Ruedy, and D. Zyss, 2019: Improvements in the GISTEMP uncertainty model. *J. Geophys. Res. Atmos.*, **124**, 6307–6326, https://doi.org/10.1029/2018JD029522.

Levang, S. J., and R. W. Schmitt, 2020: What causes the AMOC to weaken in CMIP5? *J. Climate*, **33**, 1535–1545, https://doi.org/10.1175/JCLI-D-19-0547.1.

Lindsay, R., M. Wensnahan, A. Schweiger, and J. Zhang, 2014: Evaluation of seven different atmospheric reanalysis products in the Arctic. *J. Climate*, **27**, 2588–2606, https://doi.org/10.1175/JCLI-D-13-00014.1.

Liu, J., Z. Zhang, Y. Hu, L. Chen, Y. Dai, and X. Ren, 2008: Assessment of surface air temperature over the Arctic Ocean in reanalysis and IPCC AR4 model simulations with IABP/POLES observations. *J. Geophys. Res.*, **113**, D10105, https://doi.org/10.1029/2007JD009380.

Mahajan, S., R. Zhang, and T. L. Delworth, 2011: Impact of the Atlantic meridional overturning circulation (AMOC) on Arctic surface air temperature and sea ice variability. *J. Climate*, **24**, 6573–6581, https://doi.org/10.1175/2011JCLI4002.1.

Mahlstein, I., and R. Knutti, 2011: Ocean heat transport as a cause for model uncertainty in projected Arctic warming. *J. Climate*, **24**, 1451–1460, https://doi.org/10.1175/2010JCLI3713.1.

Mouginot, J., and Coauthors, 2019: Forty-six years of Greenland Ice Sheet mass balance from 1972 to 2018. *Proc. Natl. Acad. Sci. USA*, **116**, 9239–9244, https://doi.org/10.1073/pnas.1904242116.

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

O'Neill, B. C., and Coauthors, 2016: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.*, **9**, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016.

Overland, J. E., M. Wang, N. A. Bond, J. E. Walsh, V. M. Kattsov, and W. L. Chapman, 2011: Considerations in the selection of global climate models for regional climate projections: The Arctic as a case study. *J. Climate*, **24**, 1583–1597, https://doi.org/10.1175/2010JCLI3462.1.

Pattyn, F., and Coauthors, 2018: The Greenland and Antarctic ice sheets under 1.5°C global warming. *Nat. Climate Change*, **8**, 1053–1061, https://doi.org/10.1038/s41558-018-0305-8.

Pithan, F., and T. Mauritsen, 2014: Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nat. Geosci.*, **7**, 181–184, https://doi.org/10.1038/ngeo2071.

Portnov, A., S. Vadakkepuliyambatta, J. Mienert, and A. Hubbard, 2016: Ice-sheet-driven methane storage and release in the Arctic. *Nat. Commun.*, **7**, 10314, https://doi.org/10.1038/ncomms10314.

Reintges, A., T. Martin, M. Latif, and N. S. Keenlyside, 2017: Uncertainty in twenty-first century projections of the Atlantic meridional overturning circulation in CMIP3 and CMIP5 models. *Climate Dyn.*, **49**, 1495–1511, https://doi.org/10.1007/s00382-016-3180-x.

Richter-Menge, J., and Coauthors, 2020: The Arctic [in "State of the Climate on 2019"]. *Bull. Amer. Meteor. Soc.*, **101** (8), S239–S286, https://doi.org/10.1175/BAMS-D-20-0086.1.

Rohde, R. A., and Z. Hausfather, 2020: The Berkeley Earth land/ocean temperature record. *Earth Syst. Sci. Data*, **12**, 3469–3479, https://doi.org/10.5194/essd-12-3469-2020.

——, and Coauthors, 2013: Berkeley Earth temperature averaging process. *Geoinfor. Geostat. An Overview*, **1**, 20–100, https://doi.org/10.4172/2327-4581.1000103.

Screen, J. A., 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 Coauthors, 2018: Consistency and discrepancy in the atmospheric response to Arctic sea-ice loss across climate models. *Nat. Geosci.*, **11**, 155–163, https://doi.org/10.1038/s41561-018-0059-y.

Senftleben, D., A. Lauer, and A. Karpechko, 2020: Constraining uncertainties in CMIP5 projections of September Arctic sea ice extent with observations. *J. Climate*, **33**, 1487–1503, https://doi.org/10.1175/JCLI-D-19-0075.1.

Serreze, M. C., and J. A. Francis, 2006: The Arctic amplification debate. *Climatic Change*, **76**, 241–264, https://doi.org/10.1007/s10584-005-9017-y.

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.

Stroeve, J., and D. Notz, 2018: Changing state of Arctic sea ice across all seasons. *Environ. Res. Lett.*, **13**, 103001, https://doi.org/10.1088/1748-9326/aade56.

Stuecker, M. F., and Coauthors, 2018: Polar amplification dominated by local forcing and feedbacks. *Nat. Climate Change*, **8**, 1076–1081, https://doi.org/10.1038/s41558-018-0339-y.

Tao, X., J. E. Walsh, and W. L. Chapman, 1996: An assessment of global climate model simulations of Arctic air temperatures. *J. Climate*, **9**, 1060–1076, https://doi.org/10.1175/1520-0442(1996)009<1060:AAOGCM>2.0.CO;2.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, https://doi.org/10.1029/2000JD900719.

——, R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Tokarska, K. B., M. B. Stolpe, S. Sippel, E. M. Fischer, C. J. Smith, F. Lehner, and R. Knutti, 2020: Past warming trend constrains

future warming in CMIP6 models. *Sci. Adv.*, **6**, eaaz9549, https://doi.org/10.1126/sciadv.aaz9549.

Wang, C., R. M. Graham, K. Wang, S. Gerland, and M. A. Granskog, 2019: Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: Effects on sea ice thermodynamics and evolution. *Cryosphere*, **13**, 1661–1679, https://doi.org/10.5194/tc-13-1661-2019.

Wang, S., D. Nath, W. Chen, and T. J. Ma, 2020: CMIP5 model simulations of warm Arctic–cold Eurasia pattern in winter surface air temperature anomalies. *Climate Dyn.*, **54**, 4499–4513, https://doi.org/10.1007/s00382-020-05241-2.

Wei, T., J. Li, X. Y. Rong, W. J. Dong, B. Y. Wu, and M. H. Ding, 2018: Arctic climate changes based on historical simulations (1900–2013) with the CAMS-CSM. *J. Meteor. Res.*, **32**, 881–895, https://doi.org/10.1007/s13351-018-7188-5.

Woods, C., and R. Caballero, 2016: The role of moist intrusions in winter Arctic warming and sea ice decline. *J. Climate*, **29**, 4473–4485, https://doi.org/10.1175/JCLI-D-15-0773.1.

Wu, F. Y., Q. L. You, Z. Y. Zhang, and L. Zhang, 2020: Changes and uncertainties of surface mean temperature over China under global warming of 1.5 and 2°C. *Int. J. Climatol.*, **41**, E410–E427, https://doi.org/10.1002/JOC.6694.

Wu, L. P., X. Y. Yang, and J. Y. Hu, 2019: Assessment of Arctic sea ice simulations in CMIP5 models using a synthetical skill scoring method. *Acta Oceanol. Sin.*, **38**, 48–58, https://doi.org/10.1007/s13131-019-1474-0.

Wyser, K., E. Kjellstrom, T. Koenigk, H. Martins, and R. Doscher, 2020: Warmer climate projections in EC-Earth3-Veg: The role of changes in the greenhouse gas concentrations from CMIP5 to CMIP6. *Environ. Res. Lett.*, **15**, 054020, https://doi.org/10.1088/1748-9326/ab81c2.

Xie, Y., Y. Z. Liu, and J. P. Huang, 2016: Overestimated Arctic warming and underestimated Eurasia mid-latitude warming in CMIP5 simulations. *Int. J. Climatol.*, **36**, 4475–4487, https://doi.org/10.1002/joc.4644.

You, Q., and Coauthors, 2021: Warming amplification over the Arctic Pole and Third Pole: Trends, mechanisms and consequences. *Earth-Sci. Rev.*, **217**, 103625, https://doi.org/10.1016/j.earscirev.2021.103625.

Zelinka, M. D., and Coauthors, 2020: Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.*, **47**, e2019GL085782, https://doi.org/10.1029/2019GL085782.

Zhou, T., and R. Yu, 2006: Twentieth-century surface air temperature over China and the globe simulated by coupled climate models. *J. Climate*, **19**, 5843–5858, https://doi.org/10.1175/JCLI3952.1.

——, L. Zou, and X. Chen, 2019: Commentary on the Coupled Model Intercomparison Project Phase 6 (CMIP6). *Climate Change Res.*, **15**, 445–456, https://doi.org/10.12006/J.ISSN.1673-1719.2019.193.

Zhou, X., H. Matthes, A. Rinke, B. Huang, K. Yang, and K. Dethloff, 2019: Simulating Arctic 2-m air temperature and its linear trends using the HIRHAM5 regional climate model. *Atmos. Res.*, **217**, 137–149, https://doi.org/10.1016/j.atmosres.2018.10.022.