



**Calhoun: The NPS Institutional Archive**  
**DSpace Repository**

---

Faculty and Researchers

Faculty and Researchers' Publications

---

2021-08

# A Spatial Evaluation of Arctic Sea Ice and Regional Limitations in CMIP6 Historical Simulations

Watts, Matthew; Maslowski, Wiewslaw; Lee, Younjoo J.;  
Kinney, Jaclyn Clement; Osinski, Robert

AMS

---

Watts, Matthew, et al. "A spatial evaluation of Arctic sea ice and regional limitations in CMIP6 historical simulations." *Journal of Climate* 34.15 (2021): 6399-6420.  
<http://hdl.handle.net/10945/69439>

---

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States

*Downloaded from NPS Archive: Calhoun*



Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

**Dudley Knox Library / Naval Postgraduate School**  
**411 Dyer Road / 1 University Circle**  
**Monterey, California USA 93943**

<http://www.nps.edu/library>

## A Spatial Evaluation of Arctic Sea Ice and Regional Limitations in CMIP6 Historical Simulations

MATTHEW WATTS,<sup>a</sup> WIESLAW MASLOWSKI,<sup>a</sup> YOUNJOO J. LEE,<sup>a</sup> JACLYN CLEMENT KINNEY,<sup>a</sup> AND ROBERT OSINSKI<sup>b</sup>

<sup>a</sup> *Department of Oceanography, Naval Postgraduate School, Monterey, California*

<sup>b</sup> *Institute of Oceanology of Polish Academy of Sciences, Sopot, Poland*

(Manuscript received 29 June 2020, in final form 28 April 2021)

**ABSTRACT:** The Arctic sea ice response to a warming climate is assessed in a subset of models participating in phase 6 of the Coupled Model Intercomparison Project (CMIP6), using several metrics in comparison with satellite observations and results from the Pan-Arctic Ice Ocean Modeling and Assimilation System and the Regional Arctic System Model. Our study examines the historical representation of sea ice extent, volume, and thickness using spatial analysis metrics, such as the integrated ice edge error, Brier score, and spatial probability score. We find that the CMIP6 multimodel mean captures the mean annual cycle and 1979–2014 sea ice trends remarkably well. However, individual models experience a wide range of uncertainty in the spatial distribution of sea ice when compared against satellite measurements and reanalysis data. Our metrics expose common and individual regional model biases, which sea ice temporal analyses alone do not capture. We identify large ice edge and ice thickness errors in Arctic subregions, implying possible model specific limitations in or lack of representation of some key physical processes. We postulate that many of them could be related to the oceanic forcing, especially in the marginal and shelf seas, where seasonal sea ice changes are not adequately simulated. We therefore conclude that an individual model's ability to represent the observed/reanalysis spatial distribution still remains a challenge. We propose the spatial analysis metrics as useful tools to diagnose model limitations, narrow down possible processes affecting them, and guide future model improvements critical to the representation and projections of Arctic climate change.

**KEYWORDS:** Arctic; Sea ice; Climate models; Error analysis; Model comparison; Regional models


### 1. Introduction


The Arctic is warming at twice the rate of the rest of the planet, evidenced by rising surface air temperatures in response to greenhouse gases (Serreze et al. 2009; Serreze and Barry 2011; Taylor et al. 2013; IPCC 2019). One of the most striking reflections of this Arctic amplification (Serreze and Francis 2006) is the accelerated decrease in sea ice extent (SIE; Meier et al. 2017) observed for each month of the year over the satellite record since 1978 (Serreze and Barry 2011; Stroeve and Notz 2018). Changes in the sea ice cover alter the surface albedo, the upper ocean heat content, and thus the surface energy budget of the Arctic Ocean (Jackson et al. 2011; Timmermans et al. 2018). In addition, a diminishing Arctic sea ice cover increases the air–sea exchange of momentum, surface buoyancy flux, and freshwater content (Parkinson et al. 1987; Rampal et al. 2011; Proshutinsky et al. 2019). Hence, a better understanding of the sea ice reduction is needed to improve climate predictions and projections.

The primary objective of this study is to assess and guide improvements of outstanding pan-Arctic as well as regional

limitations in historical simulations of sea ice, by employing a combination of common and new metrics on a subset of state-of-the-art Earth system models (ESMs) participating in CMIP6 (Eyring et al. 2016). This is motivated in part by the need to better understand the complex operation of the Earth system under climate forcing and in part to convey confidence in model skill to project the future. The latter is based on the argument that a model's ability to simulate “known” mean climate state, trends, variability, and extremes raises confidence in its projections (Randall et al. 2007; Massonnet et al. 2012). However, past model performance of particular observables (e.g., pan-Arctic SIE) alone is not sufficient to describe the quality of a model's future projection, due in large part to internal variability, observational uncertainty, and model tuning (Notz 2015). Internal climate variability itself allows for a range of possible outcomes of Arctic sea ice states, of which the observed state is but one realization (Notz 2015; England et al. 2019).

While modern ESMs generally capture much of the physics and the downward trends of the observed Arctic SIE, they have so far underestimated its acceleration in response to greenhouse gas forcing and increasing global surface air temperature (e.g., Winton 2011; Massonnet et al. 2012; Stroeve and Notz 2015; Rosenblum and Eisenman 2017; Notz and SIMIP Community 2020). In addition, according to observational and model reconstructed estimates, the negative trend in sea ice thickness (SIT) and volume (SIV) has been even stronger than that in SIE (Kwok and Rothrock 2009; Schweiger et al. 2011; Maslowski et al. 2012; Stroeve et al. 2014). This aspect alone corroborates the need for observationally constrained metrics

 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-20-0491.1>.

Corresponding author: Matthew Watts, [matthew.watts@nps.edu](mailto:matthew.watts@nps.edu)

DOI: 10.1175/JCLI-D-20-0491.1

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

of SIT and SIV for model evaluation as they allow for additional insights into regional and seasonal biases and overall quality of sea ice simulations. At the same time, observations of sea ice from passive microwave satellites contain internal and algorithm uncertainties (Screen 2011; Eisenman et al. 2014; Ivanova et al. 2014; Meier et al. 2014), while pan-Arctic and long-term estimates of SIT and SIV are not readily available from satellites as they cover relatively short time periods. In particular, uncertainty in satellite-derived estimates of pan-Arctic SIT distribution and summer sea ice concentration (SIC) pose considerable challenges (Zygmuntowska et al. 2013; Kwok 2018).

In this study, we expand on the published CMIP6 Arctic sea ice analyses (Shu et al. 2020; Notz and SIMIP Community 2020; Shen et al. 2021) to isolate specific spatial model limitations. In particular, we examine SIE, SIT, and SIV from a subset of 12 CMIP6 models, which provides a good representation of the whole set (as discussed in section 2), for the period 1979–2014. The integrated ice edge error (IIEE; Goessling et al. 2016) and spatial probability score (SPS; Goessling and Jung 2018) analyses, referred to collectively as ice edge analysis, are introduced to identify regions commonly challenging for the majority or individual CMIP6 models to accurately replicate sea ice conditions. The rest of this paper is organized as follows: in section 2 we describe data and methods; in section 3 results are presented, emphasizing the mean state and decline in pan-Arctic sea ice, the simulated SIT spatial distribution, and a regional ice edge analysis; and in section 4 are the discussion and conclusions.

## 2. Data and methods

### a. Model sea ice output

Sea ice outputs for a subset of CMIP6 models (Table 1) were retrieved from the Earth System Grid Federation repository (<https://esgf-node.llnl.gov/search/cmip6/>). Our study used the historical experiment data to evaluate the capability of the participating CMIP6 models to represent Arctic sea ice during the recent past. These experiments were initialized and forced with common time-dependent observations, including anthropogenic short-lived climate forcing, carbon emissions, land use, and greenhouse gas historical concentrations, for the time period of 1850–2014 (Notz et al. 2016; Eyring et al. 2016).

The 12 models selected for our study range from plus and minus two standard deviations of the CMIP6 multimodel mean sea ice area (SIA) and SIV, as shown in Table S3 in the online supplemental material of Notz and SIMIP Community (2020). We chose a limited number of models, representative of the whole with respect to multimodel mean and spread (see Table S1 of our paper), to highlight the range of simulated sea ice biases and limitations and a workable subset to present the utility of ice edge analysis. Additional criteria for model selection included the availability of sea ice variables for spatial analysis (e.g., SIT and SIC) and a representation of different sea ice model components used in CMIP6 simulations (e.g., NEMO-LIM, CICE, MPAS-Sea ice, GELATO). Table 1 summarizes the model metadata, climate system components with

nominal resolutions, and the model variables used for analysis. The horizontal resolution of the CMIP6 sea ice model component varies between 50 and 500 km, with the majority still using a relatively coarse resolution of  $1^\circ$  ( $\sim 100$  km). All but one CMIP6 model in our study, GFDL-ESM4, produced output from multiple ensemble members (up to 32). Unless otherwise indicated, all values presented in figures and tables show ensemble means of individual models.

In addition to our CMIP6 analysis, we analyze sea ice model output from the Regional Arctic System Model (RASM; Maslowski et al. 2012; Roberts et al. 2015; Hamman et al. 2016; Cassano et al. 2017). The forced sea ice–ocean model configuration, which we term RASM-G (Fig. S1 shows the RASM-G domain), was used for a high spatial resolution ( $\sim 9$  km) hindcast simulation, results of which are presented here. It was initialized after a 57-yr spinup and forced with the Japanese 55-year atmospheric surface reanalysis data for driving ocean–sea ice models (JRA55-do; Tsujino et al. 2018). This RASM-G hindcast simulation provides a complementary reconstruction of multidecadal sea ice conditions for the period of 1980–2014, in addition to the remotely sensed observations and Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS; Zhang and Rothrock 2003) reanalysis data.

Due to the lack of persistent SIT observations over the Arctic, we use the PIOMAS sea ice reanalysis as a SIT “observational” proxy reference, following a number of previous studies favorably comparing PIOMAS results against thickness observations from submarines, satellites, and airborne (Zhang and Rothrock 2003; Schweiger et al. 2011; Stroeve et al. 2014). The PIOMAS version 2.1 SIV and effective SIT were retrieved from the Polar Science Center at the University of Washington (<http://psc.apl.uw.edu/research/arctic-sea-ice-volume-anomaly/data/>) in order to evaluate SIT simulations of CMIP6 models. Here we use SIV time series and monthly mean gridded effective SIT (Zhang and Rothrock 2003; Schweiger et al. 2011).

### b. Observational sea ice data

Monthly mean SIC data for the period of 1979–2014 were retrieved from the National Oceanic and Atmospheric Administration (NOAA)/National Snow and Ice Data Center (NSIDC) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Ocean and Sea Ice Satellite Application Facility (OSI-SAF). A mean observational SIC was determined by combining the National Aeronautics and Space Administration (NASA) Team (NT) and Bootstrap (BS) SIC from the Climate Data Record (CDR) of Passive Microwave SIC, version 3 (Meier et al. 2017; [ftp://sidacs.colorado.edu/pub/DATASETS/NOAA/G02202\\_V3](ftp://sidacs.colorado.edu/pub/DATASETS/NOAA/G02202_V3)) and the OSI-SAF team SIC estimate (OSI-450; Lavergne et al. 2019; [https://doi.org/10.15770/EUM\\_SAF\\_OSI\\_0008](https://doi.org/10.15770/EUM_SAF_OSI_0008)). We use this combined SIC estimate as the primary observational reference for SIE time series analysis. For the ice edge analysis, we use the merged NT/BS SIC as observational reference data. The spread in observational estimates as a result of algorithm differences can be interpreted as the observational uncertainty, or absolute uncertainty (Meier and Stewart 2019; Notz and SIMIP Community 2020).

TABLE 1. Model simulations from phase 6 of the Coupled Model Intercomparison Project (CMIP6) and the Regional Arctic System Model (RASMG) used in this analysis. The number of ensemble members is indicated in parentheses in the “Label” column. Native model grids are used unless otherwise annotated with an asterisk (\*) indicating regridded variables used for this study are sea ice area fraction (siconc), sea ice volume per area (sivol), and sea ice mass per area (simass).

Model name	Label (No. of ensemble members)	Organization	Components: Atmosphere (A)/Ocean (O)/Sea Ice (I)	Resolution (km): A/O/I	Model variable
Canadian Earth System Model (CanESM) 5	CanESM5 (25)	Canadian Centre for Climate Modelling and Analysis	Canadian Atmosphere Model (CanAM5)/Nucleus for European Modelling of the Ocean (NEMO) v3.4.1/NEMO-Louvain-la-Neuve Sea Ice Model (LIM2); DOI: 10.22033/ESGF/CMIP6.3610; v20190429	500/100/100	siconc simass
Community Earth System Model (CESM) 2	CESM2 (11)	National Center for Atmospheric Research (NCAR)	Community Atmosphere Model (CAM6)/ Parallel Ocean Program version 2 (POP2)/Los Alamos National Laboratory Sea Ice Model version 5.1 (CICE5.1); DOI: 10.22033/ESGF/CMIP6.7627; v20190308	100/100/100	siconc sivol
CESM Whole Atmosphere Community Climate Model (WACCM)	CESM2-WACCM (3)	NCAR	Whole Atmosphere Community Climate Model (WACCM6)/POP2/CICE5.1; DOI: 10.22033/ESGF/CMIP6.10071; v20190227	100/100/100	siconc sivol
National Center Meteorological Research (CNRM) Earth System Model 2	CNRM-ESM2-1 (5)	CNRM-CERFACS	Action de Recherche Petite Echelle Grande Echelle (ARPEGE 6.3)/ NEMO3.6/Global Experimental Leads and Ice for Atmosphere and Ocean (GELATO 6.1); DOI: 10.22033/ESGF/CMIP6.4068; v20181206	250/100/100	siconc sivol
Energy Exascale ESM (E3SM)	E3SM 1.0 (5)	Department of Energy E3SM Project	E3SM Atmosphere Model (EAM)/Model for Prediction Across Scales (MPAS)-Ocean/MPAS-Sea Ice; DOI: 10.22033/ESGF/CMIP6.4497; v20190926	100/50/50	siconc* simass*
Flexible Global Ocean-Atmosphere-Land System Model (FGOALS)	FGOALS-f3-L (3)	Institute of Atmospheric Physics (IAP)	FAMIL2.2/State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG)/IAP Climate Ocean Model, (LICOM3.0)/CICE4.0; DOI: 10.22033/ESGF/CMIP6.3355; v20191031	100/100/100	siconc sivol

TABLE 1. (Continued)

Model name	Label (No. of ensemble members)	Organization	Components: Atmosphere (A)/Ocean (O)/Sea Ice (I)	Resolution (km): A/O/I	Model variable
Geophysical Fluid Dynamics Laboratory (GFDL) ESM 4	GFDL-ESM4 (1)	National Oceanic and Atmospheric Administration	GFDL-AM4.1/GFDL-OM4p5 Modular Ocean Model (MOM6)/GFDL-SIM4p5 Sea Ice Simulator (SIS2.0); DOI: 10.22033/ESGF/CMIP6.8597; v20180701	100/50/50	sicone sivol
Goddard Institute of Space Studies (GISS)	GISS-E2.1H (10)	National Aeronautics and Space Administration	GISS-E2.1/Hybrid Coordinate Ocean Model (HYCOM)/GISS Sea Ice; DOI: 10.22033/ESGF/CMIP6.7128; v20191003	250/100/100	sicone* sivol
Institut Pierre Simon Laplace (IPSL) Climate Model 6	IPSL-CM6A-LR (32)	IPSL Climate Modeling Centre	LMDz/NEMO-OPA/NEMO-Louvain-la-Neuve Sea Ice Model (LIM3); DOI: 10.22033/ESGF/CMIP6.5195; v20180803	250/100/100	sicone simass
Model for Interdisciplinary Research on Climate (MIROC) 6	MIROC6 (10)	Japan Agency for Marine-Earth Science and Technology	Center for Climate System Research (CCSR) Atmospheric General Circulation model (AGCM)/CCSR Ocean Component model (COCO4.9)/COCO4.9; DOI: 10.22033/ESGF/CMIP6.5603; v20181212	250/100/100	sicone simass
Max Planck Institute (MPI) ESM	MPI-ESM1.2-HR (10)	Max Planck Institute for Meteorology	ECHAM6.3/MPIOM1.63/unnamed; DOI: 10.22033/ESGF/CMIP6.6595; v20190710	100/50/100	sicone sivol
United Kingdom ESM (UKESM) 1	UKESM1.0-LL (14)	Met Office Hadley Centre	MetUM Hadley Centre Global Environment Model (HadGEM3-GA7.1)/NEMO-HadGEM3-GO6.0/CIICE-HadGEM3-GSI8; DOI: 10.22033/ESGF/CMIP6.6113; v20200310	250/100/100	sicone sivol
Regional Arctic System Model (RASAM) Sea Ice-Ocean	RASAM-G (1)	Naval Postgraduate School	Japanese 55-year Reanalysis (JRA55-do)/POP2/CIICE6	50/9/9	sicone sivol

Monthly mean SIT estimates from *ICESat* were retrieved from NSIDC (Yi and Zwally 2009; <https://nsidc.org/data/NSIDC-0393/versions/1>) for the period 2003–08, and *CryoSat-2* from the Alfred Wegener Institute (Hendricks and Ricker 2019; <https://spaces.awi.de/display/SIRAL>) for the period 2010–14. Satellite SIT observations are available only during the colder months from October to April, which is why we limit our SIT analysis to the month of March. For *ICESat*, SIT data are available for shorter campaign periods than *CryoSat-2* and do not align seamlessly with our selected month (but *CryoSat-2* data do), so some temporal sampling bias is expected in the model comparison.

### c. Sea ice extent, volume, and thickness

We computed simulated SIE time series using the CMIP6 variables of SIC (“siconc”) and grid cell area (“areacell”) by calculating the total area of all grid cells with  $SIC \geq 15\%$ . SIE is a prevalent metric used for model comparisons and benefits from the availability of a long-term passive microwave satellite record and the reduction of uncertainties in SIC associated with the pole hole, melt ponds, thin ice, and marginal ice zones. We chose the SIE metric, as opposed to SIA, to show results consistent with the ice edge spatial analyses methods described below (section 2d). The primary shortfall of the SIE metric is that it is strongly grid-dependent (Notz 2014) as compared to SIA, and that both SIA and SIE afford only a limited two-dimensional sea ice evaluation. In determining an ensemble model mean SIE, SIE time series for each ensemble member is first calculated before averaging.

The SIV metric incorporates the vertical dimension (i.e., thickness) and it provides a more complete measure of the state and rapid change of the Arctic sea ice (Kwok and Rothrock 2009; Stroeve et al. 2012). For each model, we computed simulated ensemble-mean SIV time series using the CMIP6 variables: for example, SIV (“sivol”) or sea ice mass (“simass”) and areacell provided by the modeling groups (see Table 1). A CMIP6 multimodel mean (MM) SIE and SIV for our study subset is determined by averaging the 12 individual model ensemble means. Sea ice anomalies are calculated relative to the 1979–2014 monthly mean for the individual models and observational references.

We also examined model simulated SIT in order to assess its spatial pattern distributions. Monthly ensemble-mean SIT was calculated on the model’s native grid (where available) using the CMIP6 variables: for example, sivol or simass multiplied by density of sea ice ( $\rho_{si}$ ), per areacell (see Table 1). A SIT cutoff of 6 m, informed by Melling (2002), was applied to all models when determining correlation and root-mean-square error (RMSE) to correct for erroneous values of too thick sea ice simulated in portions of the Canadian Arctic Archipelago (CAA) by some models.

### d. Ice edge analysis

Two ice edge spatial analysis techniques were evaluated to compare the satellite-observed SIE against model simulations of SIE. First, the spatial probability score (SPS) is a probabilistic verification score for contours (Goessling and Jung 2018). For the sea ice edge contour, SPS is defined as

$$SPS = \iint_{x,y} \left\{ P[\text{sic} > 0.15]_f(x,y) - P[\text{sic} > 0.15]_o(x,y) \right\}^2 dx dy,$$

where  $P[\text{sic} > 0.15]_f$  is the ensemble probability of having  $\text{sic} > 0.15$  and  $P[\text{sic} > 0.15]_o$  is the binary field (1 or 0) representing “perfect” SIE observations. Simply put, the SPS is a summation of the areal differences between the “true” ice edge and the probabilistic modeled ice edge. The total SPS is therefore spatial integration of local areas that are both overestimated and underestimated by model SIE. Additionally, the average of the squared difference term in the SPS equation is the local Brier score (Brier 1950), and is used here in spatial maps to evaluate each grid cell skill between zero and one in representing the ice edge (cf. Wayand et al. 2019). A Brier score of zero represents a perfect prediction of SIE, and the score of one represents the alternative extreme.

The other spatial analysis technique we apply is the integrated ice edge error (IIEE) (Goessling et al. 2016). The IIEE is a special case of SPS whereby  $P[\text{sic} > 0.15]_f$  is replaced by a binary deterministic value, as is the case for models with single member simulations. While the majority of modeling centers participating in CMIP6 do provide at least a limited number of ensemble members, there is a large range (e.g., 1 to 32). It is through this lens that we evaluate the mean IIEE against the SPS. For IIEE, each individual ensemble member was first treated as a single deterministic simulation then averaged across all individual model ensembles to determine the individual model mean. Here we calculated CMIP6 and RASM-G monthly mean SPS and IIEE over several Arctic regions.

Sea ice observations and model data used for spatial analyses were linearly regridded, as needed, onto the NSIDC Sea Ice Polar Stereographic (SIPS) North 25 km  $\times$  25 km grid to allow for cell by cell comparison between the simulated and the observed values in both time and space between models. The absolute mean errors resulting from the grid interpolation are conservatively estimated at less than 4% for SIE or SIA (up to 3% for 50-km resolution GFDL-ESM4 and up to 3.6% for 100-km resolution CESM2; Fig. S2).

## 3. Results

Prior to presenting spatial evaluation results, we first introduce time series analysis of the commonly used SIE and SIV metrics. These results reveal the simulated sea ice spread of our selected CMIP6 models and set the stage for spatial analyses. For both SIE and SIV, we find that the CMIP6 multimodel mean (MM) outperforms any single model in representing the mean state and trend of the historical sea ice cover.

### a. Sea ice extent and volume

Figure 1a shows the SIE mean annual cycle for CMIP6 and RASM-G models and the combined passive microwave observations, and Table 2 summarizes the SIE 12-month running mean, standard deviation, and linear trends from 1979 to 2014. We can clearly see that GISS-E2.1H is highest biased across all

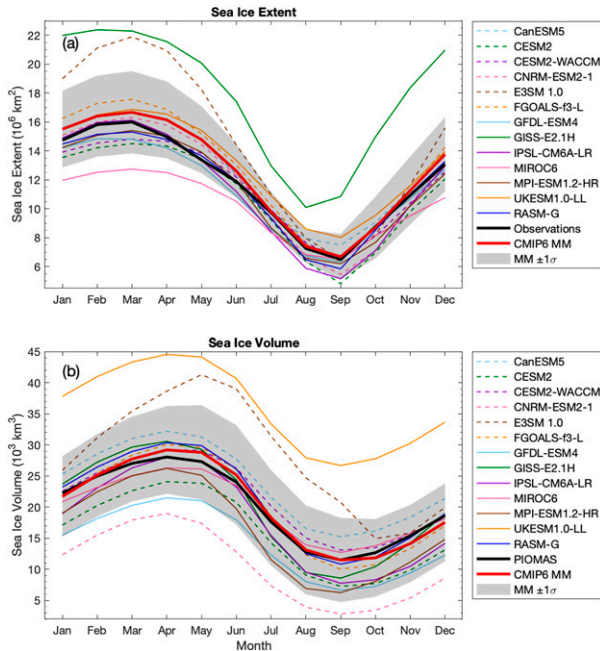


FIG. 1. Seasonal cycles of (a) SIE and (b) SIV for CMIP6 models and RASM-G with combined passive microwave observations and PIOMAS for the period 1979–2014. The CMIP6 multimodel mean and standard deviation are displayed by a red line and gray shading, respectively.

months, while E3SM 1.0 and MIROC6 both exhibit biases in winter months outside one standard deviation of the MM spread (Fig. 1a) relative to the observed mean SIE annual cycle. Additionally, all models except GISS-E2.1H exhibit a realistic seasonal cycle consisting of SIE maximum in March and minimum in September. Consistent with Shu et al. (2020), the month of March has a slightly larger spread among the CMIP6 models than September (Fig. 1a), suggesting larger winter sea ice edge variability across models.

All models simulate SIE negative trends with varying intensity (Table 2), in general agreement with the observed historical SIE decline for the 36-yr period (1979–2014). The CMIP6 MM rate of decline,  $-0.55 \times 10^6 \text{ km}^2 \text{ decade}^{-1}$ , matches closely the observed trend ( $-0.53 \times 10^6 \text{ km}^2 \text{ decade}^{-1}$ ). However, individual CMIP6 model trends are spread relatively wide, with only four of 12 CMIP6 models falling within two standard deviations of the observed trend. The discrepancy here can be mostly explained by model bias; that is, positive (negative) bias models tend to have stronger (weaker) declining trends (Table 2).

Figure 1b shows the mean SIV annual cycle for CMIP6 and RASM-G models and PIOMAS reanalysis, while Table 2 summarizes their respective SIV 12-month running means, standard deviations, and linear trends. Compared to SIE, the SIV shows a relatively larger model spread, with the largest bias in UKESM1.0-LL and CNRM-ESM2-1 (Fig. 1b). All models, except E3SM 1.0, have a realistic seasonal cycle consisting of the SIV maximum in April and the SIV minimum in September.

As in the case for SIE, all models simulate a declining SIV trend with varying intensity, which qualitatively matches the PIOMAS and satellite estimated (Kwok 2018) SIV trends. The CMIP6 MM SIV mean of  $20.36 \times 10^3 \text{ km}^3$  and trend of  $-3.02 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$  for 1979–2014 pairs very well with the PIOMAS mean of  $20.18 \times 10^3 \text{ km}^3$  and trend of  $-3.03 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$  (Table 2). Here seven of 12 CMIP6 models fall within two standard deviations of the PIOMAS SIV trend.

### b. Sea ice thickness

An accurate spatial distribution of SIT is key to estimates of SIV and it reflects the skill in simulation of local processes, coupled interactions and energy transfer between the ocean below, the sea ice, and the atmosphere above (Stroeve et al. 2014). We first assess whether the CMIP6 models accurately simulate the spatial distribution of SIT by focusing on the months of mean SIE maximum (March; Fig. 2) and minimum (September; Fig. 3) for the period 1979–2014. The mean satellite-observed ice edge, determined from the gridded NSIDC monthly SIE, is included on each SIT image. However, because of the satellite limitations in differentiating thin ice (at least up to 0.2 m) from open water (W. Meier, NSIDC, 2021, personal communication), we impose the limit  $\text{SIT} > 0.1 \text{ m}$  in order to provide a conservative and comparable estimate of the simulated ice edge.

The PIOMAS simulated monthly mean ice thickness (Figs. 2n and 3n) is often used as an historical observational “proxy” (e.g., Labe et al. 2018) with the following caveats: PIOMAS has a general tendency to underestimate SIT in regions of thick ( $>3.5 \text{ m}$ ) ice [e.g., near the Canadian Arctic Archipelago (CAA) and north of Greenland], and overestimate ice thickness in regions of thin ice. Additionally, PIOMAS has a tongue of  $\sim 2.5 \text{ m}$  ice that extends across the Arctic to the Chukchi and East Siberian Seas in March that is not depicted in situ observations (Schweiger et al. 2011; Stroeve et al. 2014). Here we include a spatial pattern correlation coefficient ( $r$ , at 99% confidence interval) and RMSE against PIOMAS on each figure. However, as was the case for CMIP5 (Stroeve et al. 2014), nearly all CMIP6 models show high correlation ( $r > 0.84$ ) with PIOMAS for March SIT (except GISS-E2.1H;  $r = 0.73$ ; Fig. 2), which makes identification of poor performers less effective, despite some models clearly underrepresenting general SIT patterns described below.

Six of 12 CMIP6 models (CanESM5, CESM2, CESM2-WACCM, FGOALS-f3-L, IPSL-CM6A-LR, and MIROC6) demonstrate a reasonable pattern of SIT relative to the PIOMAS SIT distribution reference. This means that generally they correctly locate the thickest (at least 3.5 m) Arctic sea ice along the CAA and north of Greenland, as well as thinner March sea ice located along the Eurasian shelf (Fig. 2). For September, the same six models also maintain an appreciable amount of thick ice (at least 3.0 m) along the CAA and northern Greenland. In contrast, four of the 12 CMIP6 models (CNRM-ESM2-1, GFDL-ESM4, GISS-E2.1H, and MPI-ESM1.2-HR) fail to maintain an appreciable amount of sea ice that is greater than 2.0 m at the September sea ice minimum (Fig. 3). On the other hand, UKESM1.0-LL (Figs. 2l and 3l) is laden

TABLE 2. Mean and standard deviations (s.d.) of 12-month running mean SIE and SIV and their anomaly linear trend and s.d. for CMIP6 models, RASM-G, combined passive microwave observations, and PIOMAS for the period 1979–2014 [except RASM-G for 1980–2014, marked with an asterisk (\*)].

Model	Sea ice extent				Sea ice volume			
	Mean ( $\times 10^6 \text{ km}^2$ )	s.d. ( $\times 10^6 \text{ km}^2$ )	Linear trend ( $\times 10^6 \text{ km}^2 \text{ decade}^{-1}$ )	s.d. ( $\times 10^6 \text{ km}^2 \text{ decade}^{-1}$ )	Mean ( $\times 10^3 \text{ km}^3$ )	s.d. ( $\times 10^3 \text{ km}^3$ )	Linear trend ( $\times 10^3 \text{ km}^3 \text{ decade}^{-1}$ )	s.d. ( $\times 10^3 \text{ km}^3 \text{ decade}^{-1}$ )
CanESM5	12.62	0.87	-0.84	0.06	23.82	4.15	-3.99	0.62
CESM2	10.93	0.58	-0.52	0.07	15.83	2.96	-2.78	0.48
CESM2-WACCM	11.60	0.55	-0.52	0.05	21.37	5.01	-4.82	0.37
CNRM-ESM2-1	11.85	0.39	-0.36	0.05	10.54	0.79	-0.71	0.10
E3SM 1.0	14.73	1.14	-1.06	0.11	28.30	6.04	-5.66	0.55
FGOALS-F3-L	12.88	0.38	-0.36	0.05	20.25	2.20	-2.03	0.37
GFDL-ESM4	11.32	0.46	-0.36	0.05	14.19	2.09	-1.86	0.34
GISS-E2.1H	17.83	0.74	-0.70	0.08	20.14	2.26	-2.17	0.12
IPSL-CM6A-LR	11.38	0.55	-0.53	0.07	17.72	2.84	-2.75	0.21
MIROC6	10.18	0.34	-0.31	0.04	19.80	2.48	-2.30	1.68
MPI-ESM1.2HR	11.45	0.46	-0.44	0.06	16.38	2.29	-2.17	0.20
UKESM1.0-LL	13.07	0.65	-0.61	0.06	35.99	5.32	-5.02	2.19
CMIP6 MM	12.49	2.06	-0.55	0.06	20.36	3.16	-3.02	0.60
RASM-G*	11.65	0.47	-0.44	0.04	21.03	2.64	-2.52	0.36
Observations	11.96	0.58	-0.53	0.04	—	—	—	—
PIOMAS	—	—	—	—	20.18	3.31	-3.03	0.51



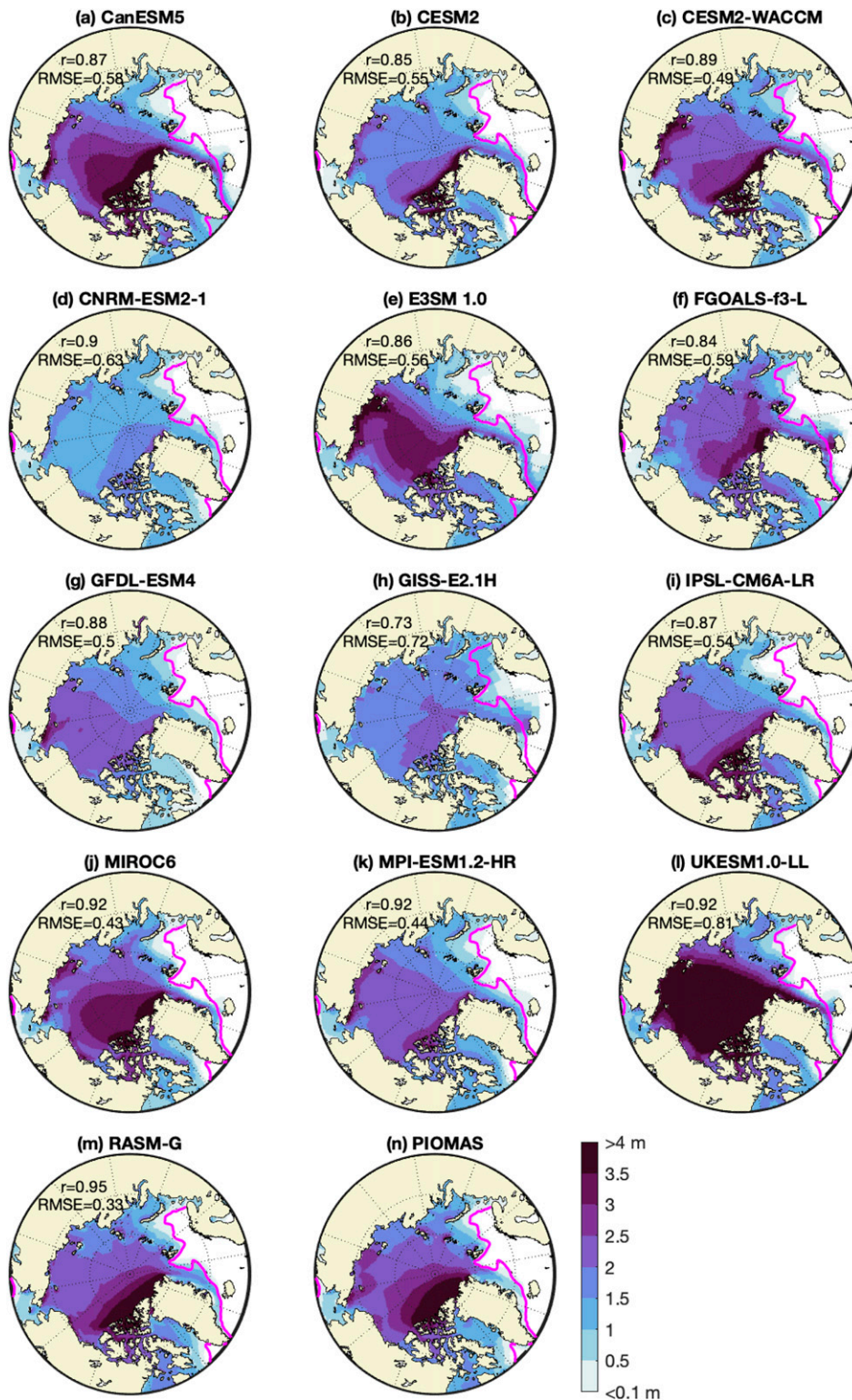


FIG. 2. March-mean sea ice thickness distribution for (a)–(l) CMIP6 models (1979–2014), (m) the RASM-G simulation (1980–2014), and (n) PIOMAS (1979–2014). Magenta contours indicate the averaged March NSIDC sea ice edge for the same period. Spatial pattern correlation coefficients ( $r$ ) and root-mean-square error (RMSE) for individual models against PIOMAS reanalysis are included in the upper-left corner of each panel.

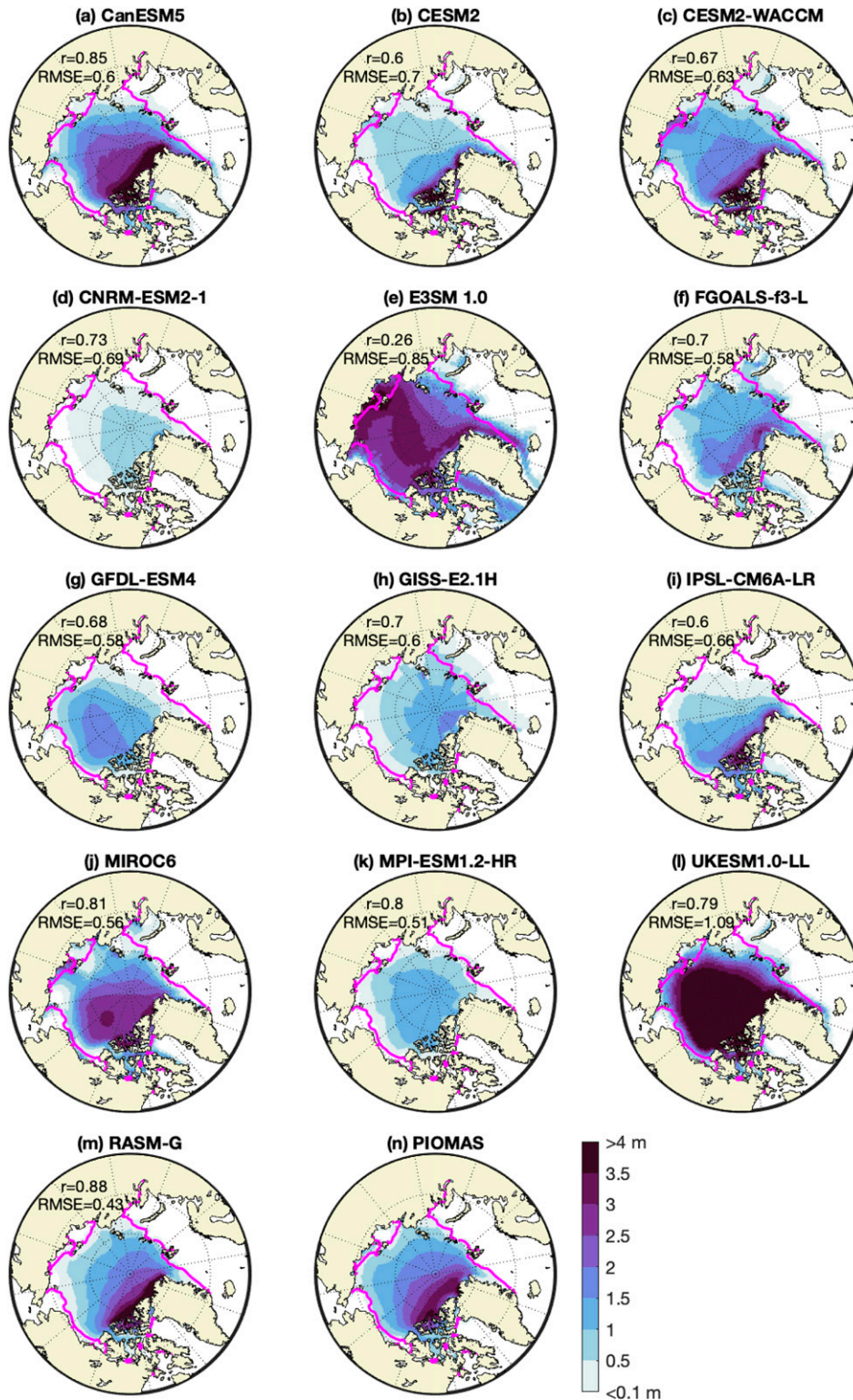


FIG. 3. September-mean sea ice thickness distribution for (a)–(l) CMIP6 models (1979–2014), (m) the RASM-G simulation (1980–2014), and (n) PIOMAS (1979–2014). Magenta contours indicate the averaged September NSIDC sea ice edge for the same period. Spatial pattern correlation coefficients ( $r$ ) and root-mean-square error (RMSE) for individual models against PIOMAS reanalysis are included in the upper-left corner of each panel.

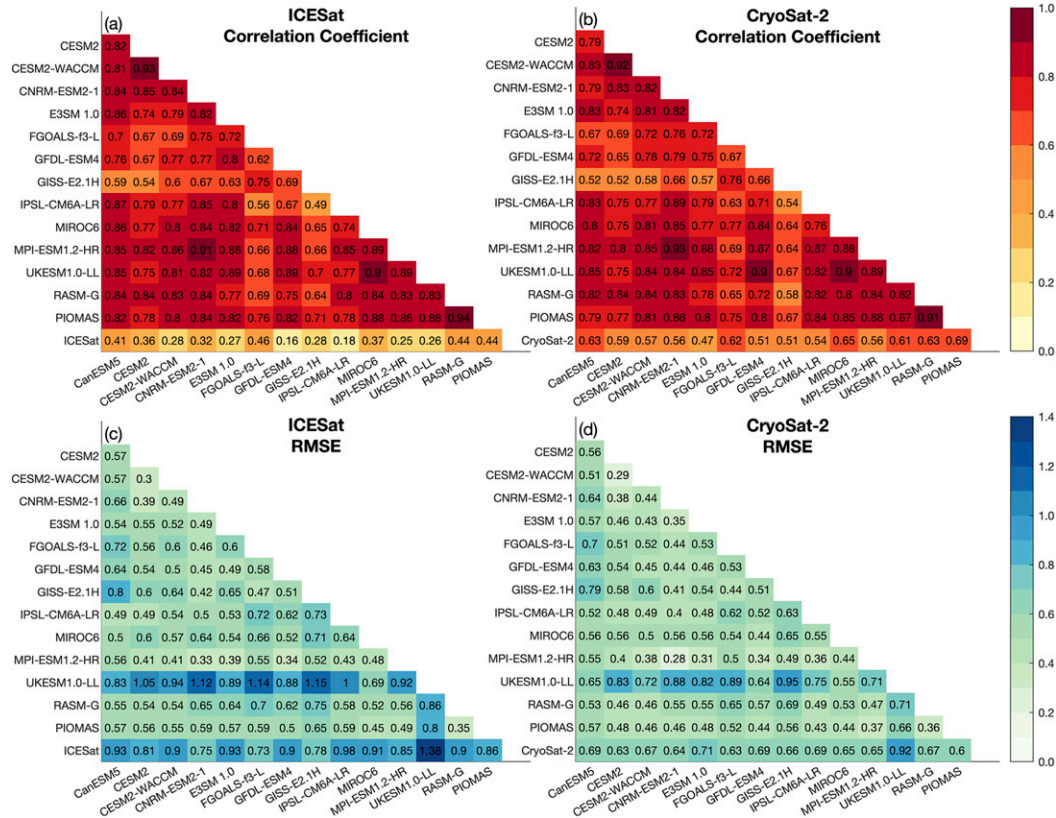


FIG. 4. (a),(c) Spatial pattern correlations and (b),(d) root-mean-square error (RMSE) of March sea ice thickness between CMIP6 models, PIOMAS and RSM-G with *ICESat* for the period of 2003–08 and *CryoSat-2* for the period of 2011–14, respectively [following Stroeve et al. (2014)].

with sea ice greater than 3.5 m throughout the analysis period and it covers a much larger area.

Turning toward the period of satellite-derived SIT observations and following Stroeve et al. (2014), we show spatial pattern correlations and RMSE between models, *ICESat*, and *CryoSat-2* in Fig. 4, and difference plots in Figs. 5 and 6. For the period 2003–08, models show low correlations against *ICESat* (all of which are significant at the 99% confidence interval) ranging from  $r = 0.18$  to 0.46 and RMSE from 0.73 to 1.38 m (Figs. 4a,c). Overall, models tend to underestimate SIT along the CAA and north and east of Greenland, and overestimate SIT over the Beaufort and Chukchi Seas and along the Siberian coast (Fig. 5). Nine of 12 CMIP6 models exhibit overestimation of SIT throughout the western Arctic (Fig. 5), which may be in part the result of the timing of *ICESat* collected observations (i.e., late February through early March) and the historically low sea ice cover during the collection years (Stroeve et al. 2014). In comparison against *CryoSat-2* for the later period 2011–14, all models show larger correlation ranging from  $r = 0.47$  to 0.65 and smaller RMSE from 0.63 to 0.92 m than against *ICESat* (Figs. 4b,d). Regional SIT differences between models and *CryoSat-2* are consistent with those against *ICESat* but are less pronounced (Fig. 6).

The RASM-G simulated SIT agrees well with PIOMAS (Figs. 2m and 3m), and *ICESat* and *CryoSat-2* (Figs. 5 and 6,

respectively). In comparison with PIOMAS, RASM-G mean SIT along the Eurasian shelf is thinner for both March and September. Additionally, RASM-G maintains a larger area of thick (>3.5 m) September sea ice along and within the CAA and north of Greenland. Evaluated against *ICESat* and *CryoSat-2* (Fig. 4), the RASM-G spatial pattern (correlation and RMSE) is highly correlated with PIOMAS reanalysis and is similar to the better-performing CMIP6 models. Overall, we find that the RASM-G simulated SIT distribution represents a comparable skill to PIOMAS.

### c. Sea ice edge

The final metrics we present for CMIP6 sea ice analysis are the SPS and the IIEE. For our subset of CMIP6 models, all but GFDL-ESM4 includes multiple ensemble members, and therefore we present the majority of our main text findings and figures from the SPS analysis. However, we show that IIEE is also an appropriate technique to compare the simulated ice edge across a spectrum of ensemble model classes (i.e., single through small to large ensemble models). Additional IIEE results are shown in the online supplemental material.

First, we analyze the SPS for the whole pan-Arctic (Fig. 7), defined here as all areas within the NSIDC SIPS North 25 km × 25 km grid with the exception of Hudson Bay, and

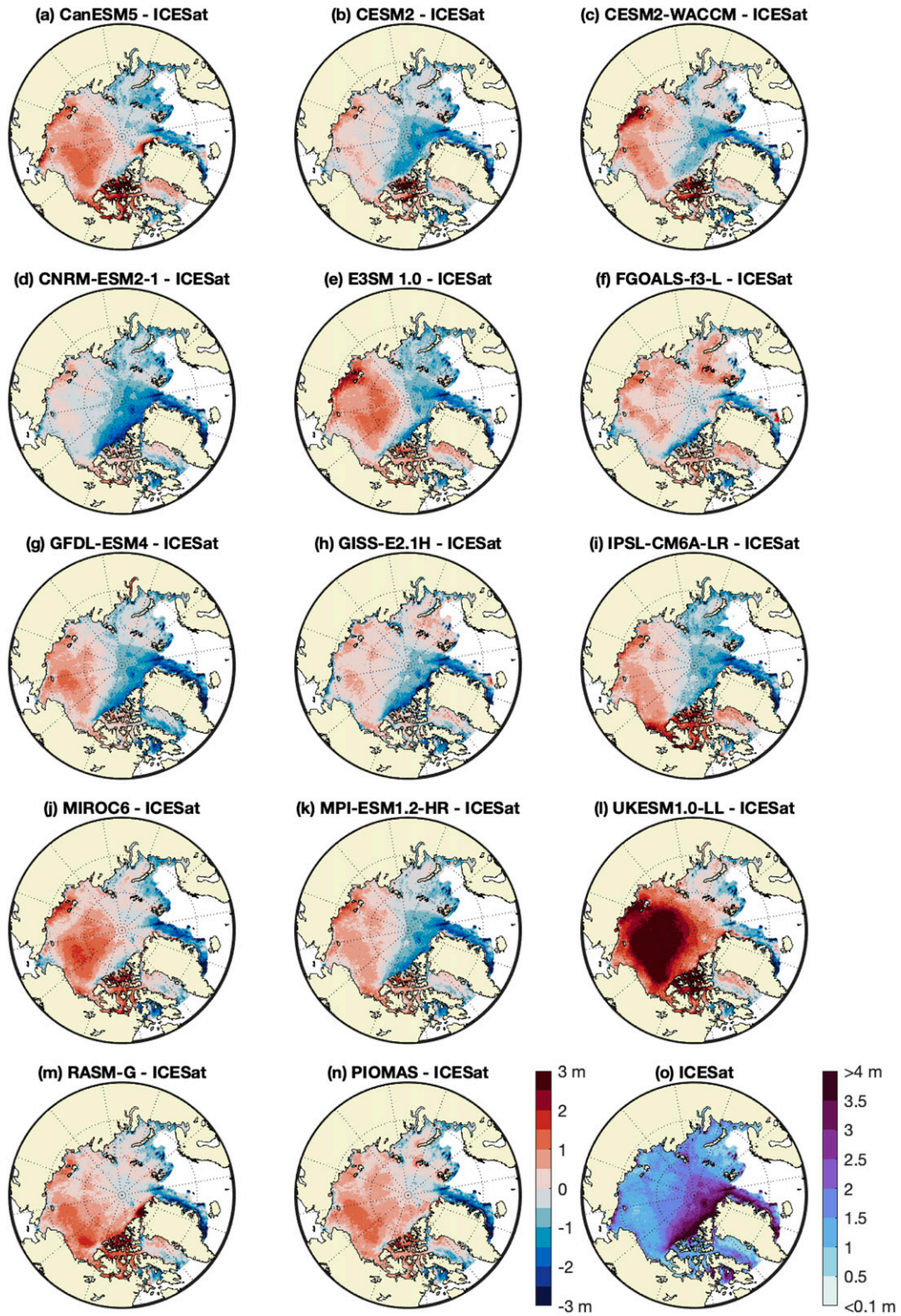


FIG. 5. March sea ice thickness differences for the period 2003–08 between (a)–(l) CMIP6 models, (m) RASM-G, (n) PIOMAS and ICESat. (o) ICESat data.

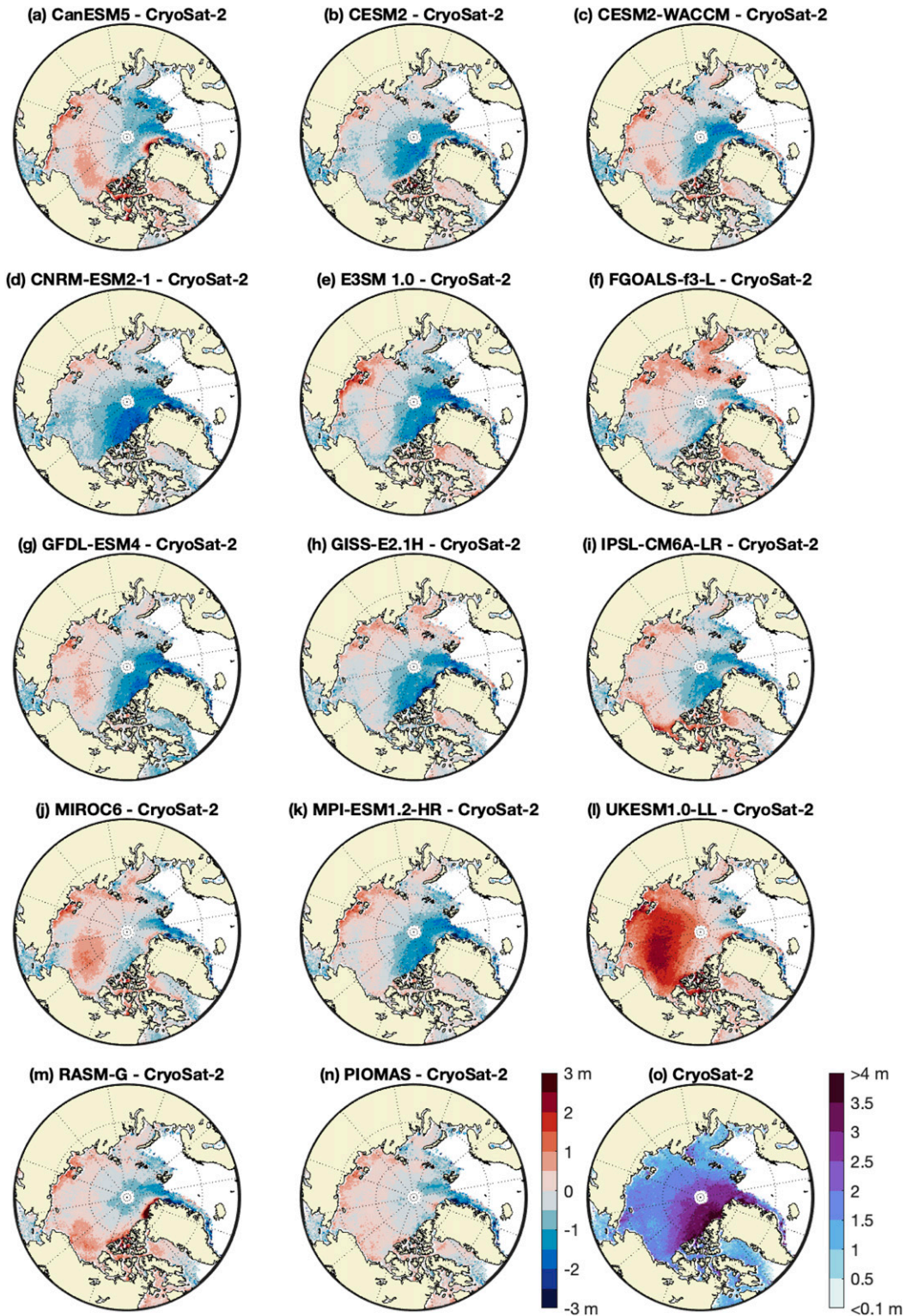


FIG. 6. March sea ice thickness differences for the period 2011–14 between (a)–(l) CMIP6 models, (m) RASM-G, (n) PIOMAS and *CryoSat-2*. (o) *CryoSat-2* data.

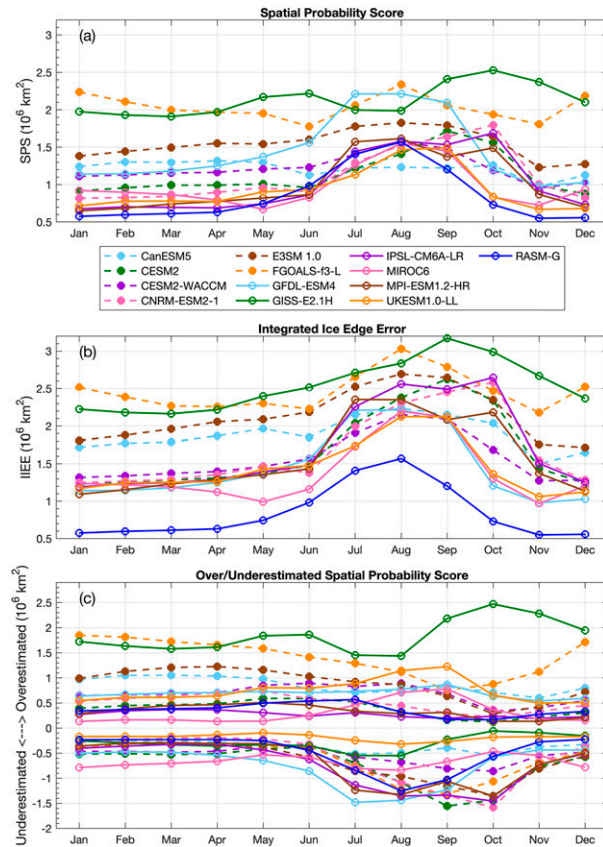


FIG. 7. Pan-Arctic monthly mean (a) SPS and (b) IIEE for the period 1980–2014 for CMIP6 models and RASM-G. (c) Pan-Arctic monthly mean SPS subdivided into two components: overestimation (positive) and underestimation (negative).

for the 11 subregions (Figs. 8 and 9). Second, we apply the SPS analysis and examine the local Brier score in order to identify regions where CMIP6 models commonly have difficulty in reproducing sea ice coverage skillfully (Figs. 10 and 11). We restrict our SPS analysis to the long-term mean errors, avoiding interannual comparisons, of which no CMIP6 models from our subset show predictive SIE skill (Fig. S3).

The resulting pan-Arctic SPS exhibits a strong mean annual cycle (Fig. 7a). Over the 35-yr record (1980–2014), most CMIP6 models have a relatively steady SPS during the colder months of December through April, with relatively better-performing models showing SPS ranges between  $0.7 \times 10^6 \text{ km}^2$  and  $1.0 \times 10^6 \text{ km}^2$  (CESM2, CNRM-ESM2-1, IPSL-CM6A-LR, MIROC6, MPI-ESM1.2-HR, and UKESM1.0-LL). All models, except CanESM5, experience increased SPS during the warmer summer/autumn months (July through October) ranging from  $0.8 \times 10^6 \text{ km}^2$  to  $2.5 \times 10^6 \text{ km}^2$ . FGOALS-f3-L and GISS-E2.1H have less dramatic seasonal cycles due to their colder season SPS remaining high, when their minimum ranges from  $1.9 \times 10^6 \text{ km}^2$  to  $2.3 \times 10^6 \text{ km}^2$ . The timing of individual model SPS peak ranges from July through October, with the majority of CMIP6 models peaking in

August (five models) or October (three models). GISS-E2.1H has the largest single month SPS, near  $2.5 \times 10^6 \text{ km}^2$  in October. The lowest single month SPS values, near  $0.7 \times 10^6 \text{ km}^2$ , occur in May (MIROC6) and November (UKESM1.0-LL).

Compared against the SPS, the pan-Arctic mean IIEE magnitude is always larger, by annual mean factor ranging between 1.2 (GISS-E2.1H) to 1.7 (IPSL-CM6A-LR), for all models with multiple ensemble members (Fig. 7b). Additionally, the shape of a model's mean annual cycle and relative performance against other models remains largely consistent between the chosen metrics (except the shape of GISS-E2.1H). Note that the single ensemble member GFDL-ESM4 SPS and IIEE are the same.

In Fig. S5, we show the individual model ensemble member's IIEE alongside their mean IIEE and SPS. This detailed look illustrates the IIEE spread about the model's mean across the whole seasonal cycle, which is not evident from the SPS alone. The individual model spread between ensemble realizations can be interpreted as the model internal variability in simulating the sea ice edge.

In Fig. 7c, the two separate components of pan-Arctic SPS, overestimation (SPS-O) and underestimation (SPS-U), are further examined. For December through April, SPS-O has a greater intermodel spread than SPS-U. Neither component of SPS has much change from month to month during this period. However, during the spring melt through autumn freeze up, individual models do show considerable temporal variability in SPS (i.e., variability in the position of the sea ice edge). Magnitudes of SPS-O in five of 12 CMIP6 models (CanESM5, CNRM-ESM2-1, GFDL-ESM4, IPSL-CM6A-LR, and MPI-ESM1.2-HR) exhibit relatively small ( $<0.4 \times 10^6 \text{ km}^2$ ) changes from June through November (Fig. 7c). However, MIROC6, UKESM1.0-LL, and GISS-E2.1H have increased values of SPS-O during the same period, suggesting that the simulated sea ice is not melting in the right regions and/or as fast as observed. On the other hand, E3SM 1.0, FGOALS-f3-L, and CESM2-WACCM models have decreased values of SPS-O. In contrast, most CMIP6 models (all except UKESM1.0-LL and CanESM5) have distinct seasonal surges of SPS-U for June through November. Four of 12 CMIP6 models have peak SPS-U in July/August and six models peak in September/October. Models with larger SPS-U early in the melt season (July/August) suggests that sea ice is removed by melting or advection faster than observed. Models with peak underestimation occurring later into the freeze up season (September/October) suggest that sea ice is not growing quick enough in their simulations.

Figure 8a shows the 35-yr mean monthly SPS-O/U for individual models and the relative contribution per Arctic subregion (defined in Fig. 8b) for March. During both winter and spring (December through May; Fig. S6), the areas that contribute to SPS are predominantly limited to the sub-Arctic seas (i.e., the Bering Sea, Nordic seas, and Baffin Bay) because the ice-covered Arctic interior is all frozen and well represented by all CMIP6 models during this period. The majority of the SPS is composed of overestimations across the Greenland and

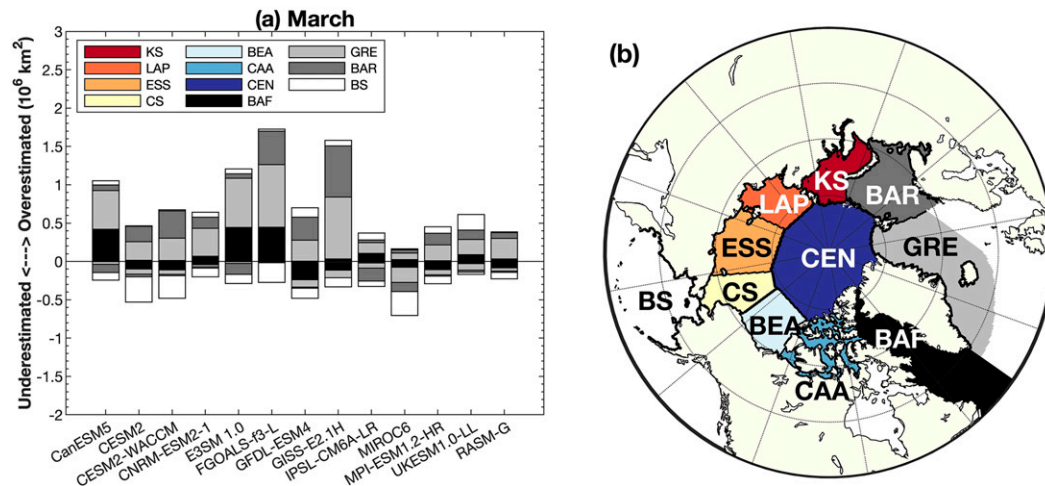


FIG. 8. (a) March monthly mean SPS overestimation and underestimation for the period 1980–2014 from CMIP6 models and RASMG and the relative contribution of the Arctic subregions. (b) Arctic regional mask defined in this study: KS = Kara Sea; LAP = Laptev Sea; ESS = East Siberian Sea; CS = Chukchi Sea; BEA = Beaufort Sea; CAA = Canadian Arctic Archipelago; CEN = Central Arctic; BAF = Baffin Bay; GRE = Greenland Sea; BAR = Barents Sea; BS = Bering Sea (adapted from Meier et al. 2007).

Barents Seas and underestimations for the Bering Sea. We use stacked SPS bar charts in Figs. 8 and 9 to identify the common regions of large SPS-O/U, as well as diagnose differences between the seasonal evolution of individual model SPS. For example, CanESM5 and GISS-E2.1H have the same region of primary overestimation over the Greenland Sea, but have different secondary error regions; for CanESM5 it is the Baffin Bay, whereas for GISS-E2.1H it is the Barents Sea (Fig. 8a).

Figure 10 displays the mean local Brier score for March and clearly depicts individual model regions contributing to the total SPS. However, because the Brier score is the squared difference, it does not distinguish between overestimation or underestimation of the ice edge. GISS-E2.1H, E3SM 1.0, FGOALS-f3-L, and CanESM5 have particularly high winter SPS values, which can be attributed to a substantial overestimation of the ice edge position in the Greenland Sea (Figs. 8a and 10). Additionally, GISS-E2.1H and FGOALS-f3-L notably overestimate the Barents Sea ice edge, and E3SM 1.0, FGOALS-f3-L, and CanESM5 overestimate the position of the Baffin Bay ice edge. On the other hand, MIROC6 is identified as having the largest SPS-U during this period, which can be attributed to a large underestimation of ice edge in both the Greenland and Bering Seas (Figs. 8a and 10).

Next, we examine the remaining, mostly warmer Arctic months of June through November (Fig. 9). Three principal regions can be identified as the largest contributors of SPS for each month. As sea ice retreats poleward, the number of regions that contribute to SPS increases and evolves in time. In June (Fig. 9a), the Greenland and Barents Seas and Baffin Bay are the regions of primary SPS (predominantly overestimation). This suggests that as the observed sea ice is retreating, the CMIP6 models tend to melt sea ice too slowly. In July (Fig. 9b), Barents Sea overestimation decreases (but still

remains a substantial error source) and the underestimation of Baffin Bay and the Kara Sea is increasing. This suggests that the simulated sea ice over the Barents Sea is catching up (i.e., accelerating modeled sea ice retreat) from a slow start of sea ice melt, and that sea ice over Baffin Bay and the Kara Sea reduces faster than the observed retreat. August is dominated by SPS-U across the East Siberian and Kara Seas, and CAA (Figs. 9c), implying that simulated sea ice along the Arctic periphery generally melts quicker than observed. In September (Figs. 9d and 11), the CAA and central Arctic Ocean ice edges are generally underestimated by CMIP6 models. On the other hand, the East Siberian Sea is nearly split among ice edge overestimation and underestimation. October ice edge errors are dominated by underestimation over the East Siberian, Kara, and Beaufort Seas (Fig. 9e), suggesting that sea ice growth is slower in a number of model simulations relative to observations. November SPS (Fig. 9f) is dominated by overestimation of the Greenland and Barents Seas, and underestimation of the Chukchi Sea. This suggests the CMIP6 models are over efficient in sea ice growth over the Greenland and Barents Seas and have slow growth over the Chukchi Sea.

Details of model specific regional SPS evaluation are summarized in Table 3, which shows the regional SPS-O/U values accumulated over the months of greatest SPS spatial variability within the Arctic, June through November. The regions of the Barents Sea and Baffin Bay account for nine of 12 CMIP6 models largest SPS-O during this 6-month period, while the Greenland Sea accounts for the other three models (CESM2, CNRM-ESM2-1, and MIROC6). The regions with the largest SPS-U are the Kara Sea (CanESM5, CESM2-WACCM, E3SM 1.0, IPSL-CM6A-LR, and UKESM1.0-LL), the East Siberian Sea (CESM2 and CNRM-ESM2-1), and Baffin Bay (GFDL-ESM4 and GISS-E2.1H). Restricting analysis in Table 3 to the extended-interior Arctic (excluding the Barents and Greenland Seas and

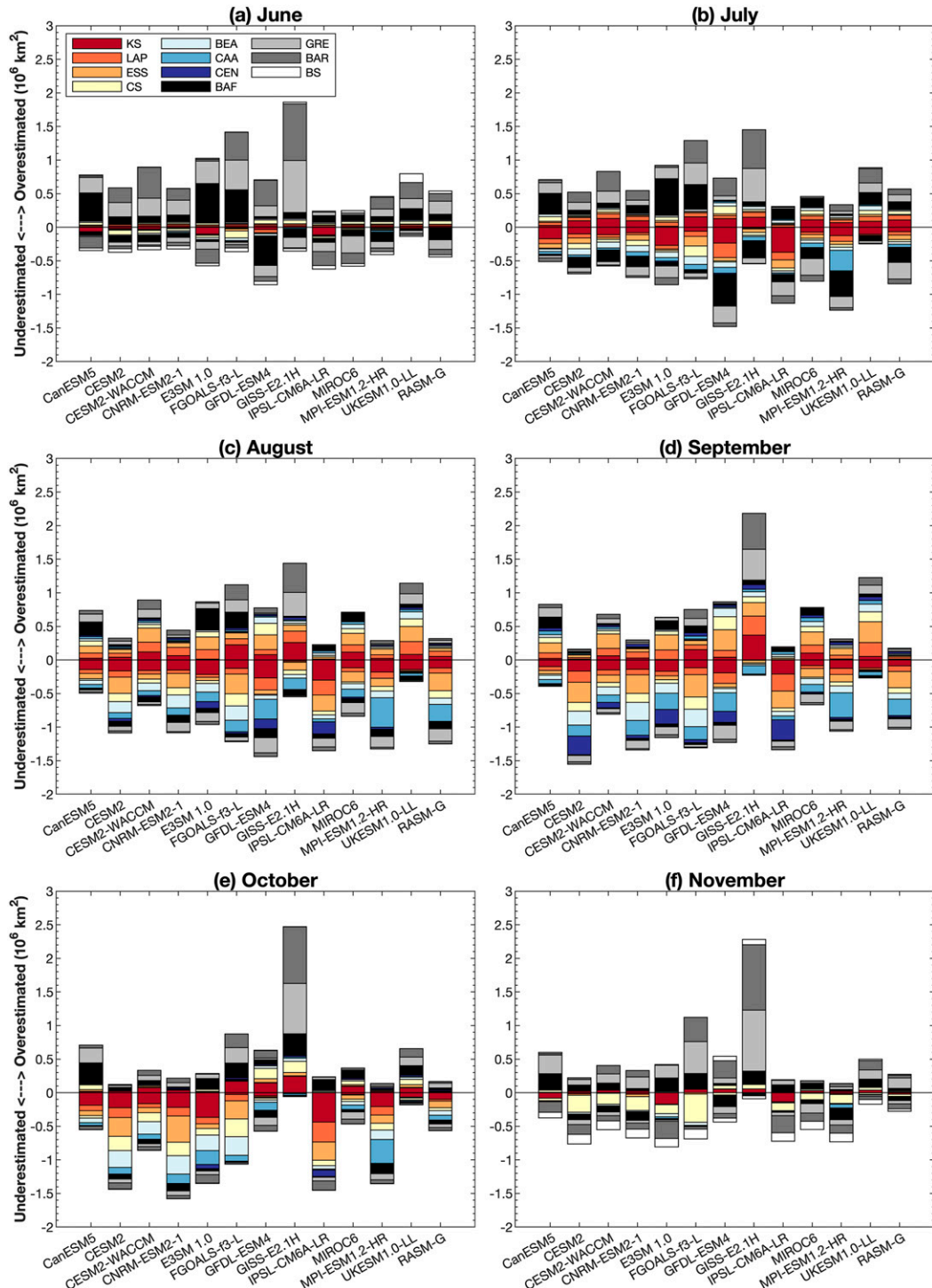


FIG. 9. Monthly mean SPS overestimation and underestimation for the period 1980–2014 from CMIP6 models and RASM-G and the relative contribution of the Arctic subregions (see Fig. 8b) for (a) June–(f) November.

Baffin Bay), we find that the Kara and East Siberian Seas are the regions with the largest SPS, either for overestimation or underestimation, for most of the CMIP6 models. Of note, four of 12 CMIP6 models (CESM2, E3SM 1.0, MIROC6, and

UKESM1.0-LL) have an inverse relationship between regions with largest SPS-O and largest SPS-U in the Kara and East Siberian Seas; that is, the Kara (East Siberian) Sea is most overestimated and the East Siberian (Kara) Sea is most



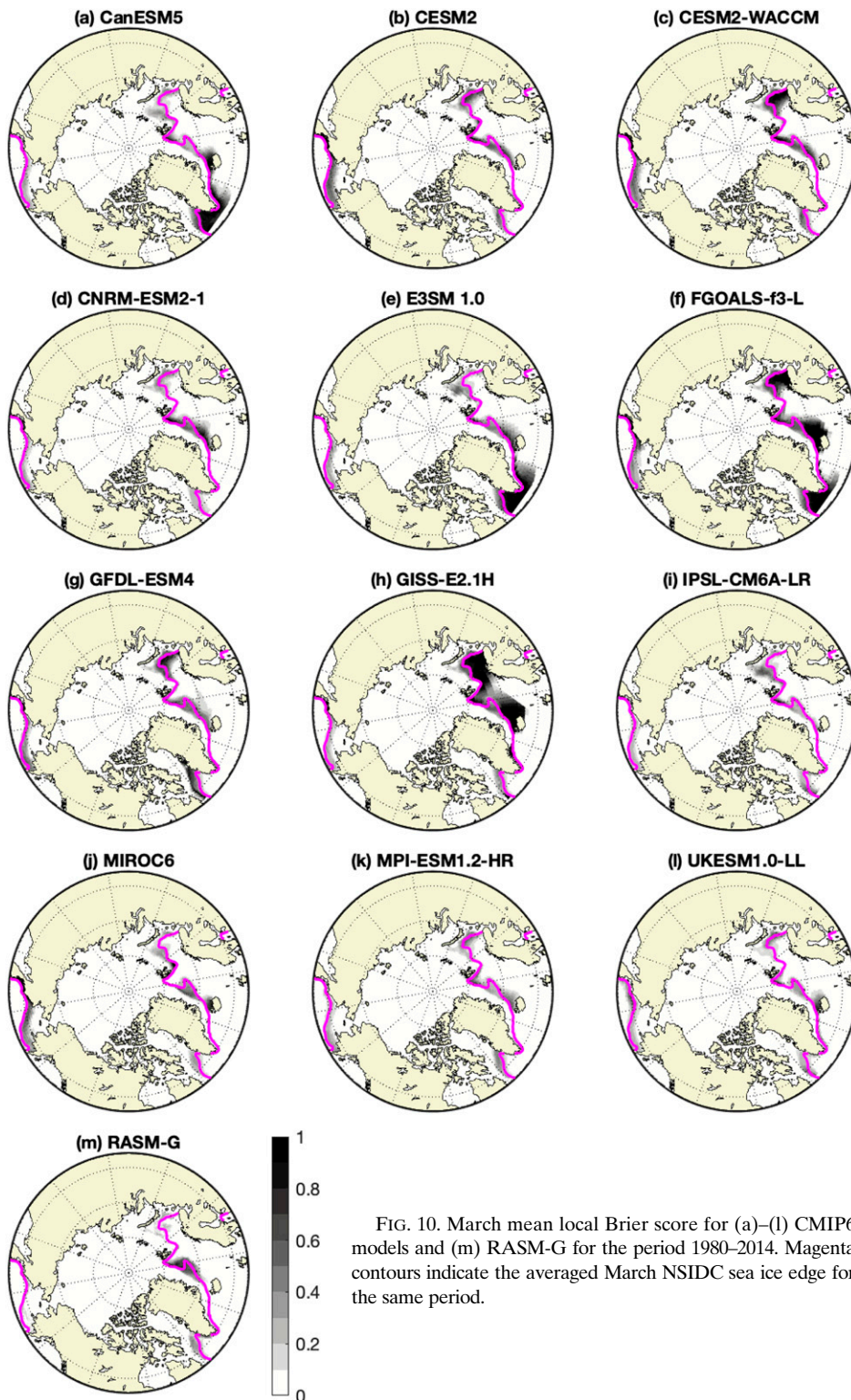


FIG. 10. March mean local Brier score for (a)–(l) CMIP6 models and (m) RASM-G for the period 1980–2014. Magenta contours indicate the averaged March NSIDC sea ice edge for the same period.

underestimated. The other regions of peak SPS are the Chukchi Sea (CanESM5 and GFDL-ESM4) and the Beaufort Sea (IPSL-CM6A-LR) for SPS-O, and CAA (GFDL-ESM4 and MPI-ESM1.2-HR) for peak SPS-U.

The RASM-G sea ice simulation shares the strong SPS and IIEE seasonality observed in the majority of CMIP6 models and has a peak in August (Fig. 7). However, compared to those models its IIEE is notably lower, by  $0.5\text{--}1.0 \times 10^6 \text{ km}^2$ ,

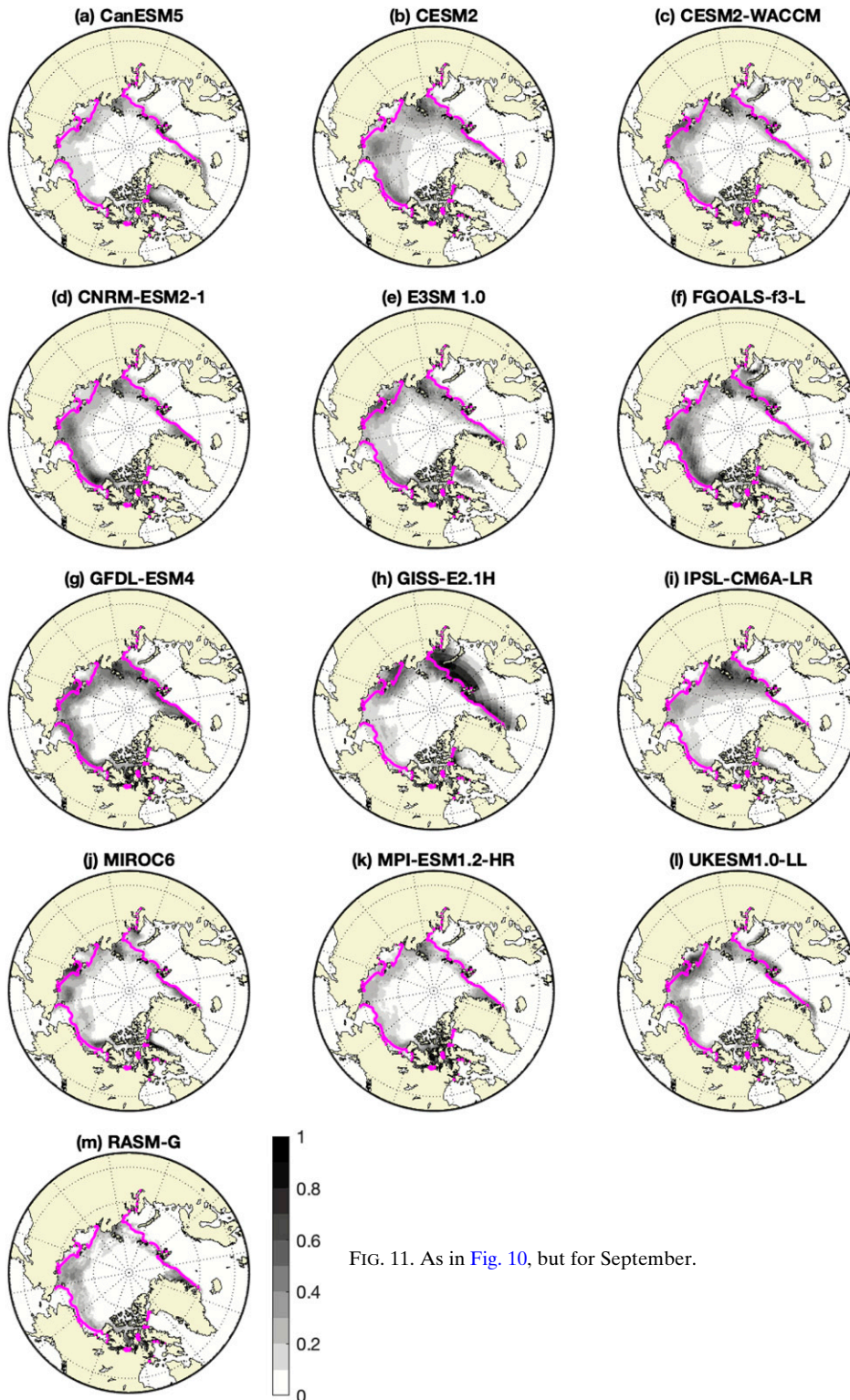


FIG. 11. As in Fig. 10, but for September.

throughout the year (Fig. 7b). The RASM-G simulation has a relatively steady SPS-O of  $0.3\text{--}0.4 \times 10^6 \text{ km}^2$  throughout the year, and a slightly smaller baseline SPS-U from December through April (Fig. 7c). The region that contributes the

greatest ice edge error for the RASM-G simulation from June through November is the Greenland Sea (Table 3). Limiting our scope to the extended-interior Arctic (defined above), we find the regions with greatest errors in RASM-G are SPS-O of

TABLE 3. Accumulated 35-yr (1980–2014) monthly mean spatial probability score ( $\times 10^6 \text{ km}^2$ ; SPS) for the months of the greatest SPS spatial variability within the Arctic, June through November. Boldface values indicate individual model maximum SPS overestimation (O) and boldface italics indicate the maximum SPS underestimation (U) for each region (see Fig. 8b); KS = Kara Sea; LAP = Laptev Sea; ESS = East Siberian Sea; CS = Chukchi Sea; BEA = Beaufort Sea; CAA = Canadian Arctic Archipelago; CEN = Central Arctic; BAF = Baffin Bay; GRE = Greenland Sea; BAR = Barents Sea; BS = Bering Sea). The largest SPS-O and SPS-U in each region are indicated by an asterisk (\*).

Region		CESM2-			GFDL-			GISS-			MIROC6	MPI-ESM1.2HR	UKESM1.0-LL	RASMG
		CanESM5	CESM2	WACCM	ESM2-1	ESM2-1	ESM4	E2.1H	IPSL-CM6A-LR	MIROC6				
KS	O	0.11	0.22	0.45	0.28	0.05	0.78	0.47	*1.10	0.02	0.48	0.17	0.35	0.22
	U	<b>0.75</b>	0.69	<b>0.61</b>	0.62	<b>1.30</b>	0.34	0.78	0.03	<b>*1.57</b>	0.32	0.67	<b>0.46</b>	0.40
LS	O	0.22	0.15	0.35	0.30	0.37	0.24	0.25	*0.55	0.07	0.31	0.17	0.50	0.20
	U	0.28	0.47	0.24	0.29	0.27	0.37	0.62	0.03	*0.90	0.15	0.41	0.07	0.25
ESS	O	0.31	0.12	0.48	0.15	0.46	0.19	*0.61	0.37	0.08	0.41	0.23	0.61	0.07
	U	0.27	<b>0.94</b>	0.30	<b>1.00</b>	0.20	*1.10	0.19	0.21	0.90	0.40	0.40	0.04	0.68
CS	O	0.36	0.04	0.18	0.06	0.26	0.03	*0.75	0.43	0.11	0.33	0.11	0.48	0.16
	U	0.15	0.87	0.50	0.76	0.39	<b>*1.30</b>	0.12	0.14	0.38	0.22	0.42	0.09	0.31
BEA	O	0.14	0.05	0.09	0.05	0.08	0.08	0.32	0.21	0.14	0.21	0.10	*0.34	0.07
	U	0.24	0.76	0.56	0.86	0.66	*0.95	0.28	0.10	0.21	0.28	0.41	0.04	0.30
CAA	O	0.11	0.06	0.09	0.04	0.04	0.09	0.07	0.10	0.11	0.12	0.01	*0.13	0.03
	U	0.21	0.39	0.30	0.57	0.68	0.55	0.80	0.41	0.14	0.37	<b>*1.55</b>	0.13	0.63
CEN	O	0.09	0.06	0.08	0.10	0.05	0.12	0.08	*0.14	0.02	0.08	0.09	0.12	0.08
	U	0.02	0.33	0.11	0.06	0.40	0.08	0.34	0.00	*0.60	0.05	0.06	0.01	0.04
BB	O	<b>1.60</b>	0.27	0.31	0.24	<b>*1.85</b>	1.60	0.13	0.73	<b>0.60</b>	<b>0.56</b>	0.08	0.68	0.25
	U	0.05	0.47	0.50	<b>0.59</b>	0.23	0.15	<b>*1.33</b>	<b>0.50</b>	0.26	0.40	0.98	0.16	0.58
GRE	O	1.14	<b>0.51</b>	0.86	<b>0.59</b>	0.89	<b>1.77</b>	<b>1.07</b>	*3.74	0.20	0.05	0.25	0.84	<b>0.59</b>
	U	0.11	0.42	0.29	0.56	0.62	0.21	*1.07	0.28	0.65	<b>1.04</b>	0.72	0.13	<b>0.80</b>
BAR	O	0.24	0.47	<b>1.15</b>	0.59	0.10	1.67	<b>1.12</b>	<b>*4.20</b>	0.05	0.16	<b>0.44</b>	<b>0.98</b>	0.31
	U	0.46	0.37	0.21	0.25	0.79	0.16	0.39	0.03	*0.83	0.52	0.32	0.12	0.34
BS	O	0.03	0.00	0.01	0.01	0.02	0.00	0.08	0.11	0.00	0.04	0.02	*0.17	0.06
	U	0.13	0.20	0.18	0.18	0.17	*0.20	0.12	0.09	0.18	0.16	0.18	0.09	0.07
Pan-Arctic	O	4.36	1.94	4.03	2.41	4.16	6.57	4.25	11.69	1.40	2.74	1.67	5.20	2.04
	U	2.68	5.91	3.80	5.75	5.71	5.41	6.02	1.83	6.62	3.90	6.11	1.32	4.40

$0.22 \times 10^6 \text{ km}^2$  across the Kara Sea and SPS-U of  $0.68 \times 10^6 \text{ km}^2$  across the East Siberian Sea (Table 3).

#### 4. Discussion and conclusions

In light of modest improvements of sea ice simulation in CMIP6 models over previous CMIP phases (Davy and Outten 2020; Shu et al. 2020; Notz and SIMIP Community 2020; Shen et al. 2021), we investigated how these models represent the spatial patterns of ice thickness and ice edge. While standard sea ice integrated analyses are good in identifying highly biased models, we caution against the potential for seemingly “good” models arriving at the “right” answer for the wrong reasons, especially without ensuring that the responsible physics to get the state correct is reasonably resolved and not by chance of internal variability (Stroeve and Notz 2015; Jahn et al. 2016). We argue that basic comparisons of time series of integrated SIE and SIV are not enough to qualify model performance and diagnose biases. Hence, we present spatial analysis techniques to assist.

Qualitative examination of the SIT spatial distributions showed that half of the analyzed CMIP6 models reasonably simulated March and September SIT (Figs. 2 and 3) for the period 1979–2014 against PIOMAS reanalysis. However, spatial pattern correlation and RMSE (Fig. 4) do not clearly distinguish models with poor SIT patterns or lend themselves to regional diagnostics. Because of the narrow range of values—for example, PIOMAS has ranges of  $r = 0.73$  to  $0.92$  and  $\text{RMSE} = 0.43$  to  $0.81 \text{ m}$  (Fig. 2) and *CryoSat-2* has ranges of  $r = 0.47$  to  $0.65$  and  $\text{RMSE} = 0.63$  to  $0.92 \text{ m}$  (Figs. 4b,d)—it is difficult to classify meaningful SIT differences between models.

We performed an ice edge analysis and identified several common “problem” regions, which may benefit from further regional analyses at the individual model level. Note that the analyses with SIA carry similar, if not larger, uncertainties related to coarse resolution in both models and satellite measurements, especially due to melt pond coverage during the melt season, detection of thin ice (by passive microwave), and the marginal ice zones (Ivanova et al. 2014; Yang et al. 2016; Comiso et al. 2017). To our understanding, the limitation with differentiating melt ponds and low SIC from open ocean surface by passive microwave satellites is the primary reason why the NSIDC and other centers choose SIE as their primary and less uncertain diagnostics of summertime ice cover. Given the above concerns with SIA, we find SPS and IIEE as reasonable options to quantify regional errors in the simulated sea ice cover and to narrow possible oceanic/atmospheric drivers of such errors. Hence, given the continued challenges with accurate SIA estimates the ice edge analysis appears to be a promising methodology for diagnosing model biases related to specific regional processes.

Our spatial analysis metrics identified Arctic regions with large ice edge and ice thickness errors, therefore pointing to limitations in or lack of representation of some physical processes within individual CMIP6 models. We postulate that many of them could be related to the oceanic forcing in the marginal and shelf seas. For example, during the warmer months, the SPS in the interior Arctic is generally largest over

the East Siberian and Kara Seas (Table 3), with many sea ice simulations tending to respond slowly to the seasonal changes (i.e., slow to melt or/and slow to grow). In case of errors of ice growth in these regions, we hypothesize this may be the result of models taking additional time to ventilate the excess heat accumulated in the upper ocean through the summer, to cool water down to freezing temperature in order to begin the ice growth. Another related cause of such errors might be an over estimation of the surface mixed layer depth, which would also take more time to cool down before freezing. Note that these two issues would affect a model representation of Arctic amplification hence more than just simulation of sea ice.

In the case of models that consistently underestimate the position of the sea ice edge for the Barents and Kara Seas (e.g., GFDL-ESM4 and IPSL-CM6A-LR in Figs. 3g,i), possible causes may be an overestimated transport of warm Atlantic Water across or underestimated cooling over the Barents Sea (Maslowski et al. 2012). The resulting excess of oceanic heat transport could be accelerating ice melt and delaying freeze-up over the Barents Sea and farther downstream, over the northern Kara Sea (and beyond).

During the colder months, the majority of SPS are overestimations over the Greenland and Barents Seas and underestimations over the Bering Sea. On the European side, this could again be related to variability in volume and heat fluxes from the North Atlantic Ocean and their distribution between the Labrador, Greenland, and Barents Seas, which may lead to interrelated inaccuracies in simulated sea ice over those regions. For example, if too much warm Atlantic Water is diverted from entering the Labrador or eastern Greenland Sea, and instead enters the Barents Sea, we could expect an overestimation of ice melt in the latter region with underestimation of melt in the former two regions (e.g., IPSL-CM6A-LR in Fig. 2i). Onarheim et al. (2018) found that the Barents Sea, considered a hotspot for recent climate change, contributes the largest fraction of winter sea ice loss through 2016. Possible explanations of the significant loss in the Barents Sea include a decreased sea ice import and the impact on upper ocean stratification (Lind et al. 2018) and greater control of the sea ice cover after around 2003 by Atlantic Water ocean heat anomalies across the Nordic seas and Barents Sea (Schlichtholz 2019). We found that the majority of CMIP6 models, consistent with Shu et al. (2020), underrepresent the strength of March SIC decline over the Barents Sea compared to NSIDC (not shown), which is indicative of potential model limitations in representing these dynamics.

Sea ice edge and thickness distribution are also determined by atmospheric forcing. Previous CMIP studies have shown that inaccurate SIT patterns are often associated with model deficiencies in simulating the observed Arctic sea level pressure patterns and associated geostrophic winds (Schweiger et al. 2011; Stroeve et al. 2014). Inaccurate simulations of the position and/or strength of large-scale circulation patterns, like the Beaufort high, can allow for sea ice convergence in regions not generally observed (DeRepentigny et al. 2016). Additionally, strong cyclones and anomalous circulation patterns can also contribute to anomalous SIT and SIE patterns in model simulations. For example, atmospheric heat and moisture drawn

into the high-latitudes can increase downward longwave radiation, which inhibits sea ice growth (Cai et al. 2020). Also, synoptic storms can excite inertial oscillations in sea ice (Roberts et al. 2015), which can lead to increased deformation rates and hence change in SIT.

Modeled SIT in the Bering Sea generally replicates observed ice thickness patterns well, but models tend to underestimate the SIE there. Given that sea ice in the Bering Sea melts completely every summer, most of the CMIP6 models get the SIT correct as long as they can replicate that seasonal retreat. However, first-year sea ice is more susceptible to atmospheric forcing and its variability (Rampal et al. 2011), hence the problems with simulation of the ice edge there. Model limitations in representing the Bering Sea dynamics also affect the Chukchi and East Siberian Seas downstream, as those marginal seas are tightly connected via the advection of Pacific water (Maslowski et al. 2014). The northeastward transport of warm Pacific summer water across Bering Strait into the East Siberian Sea is another process likely to be difficult to represent in the CMIP6 models with coarse horizontal resolution (e.g., the width of the Bering Strait is less than the single grid cell size for all but three models examined here). Yet the oceanic forcing of sea ice in the East Siberian Sea might play a role in its negative concentration trend, which Onarheim et al. (2018) identified as the largest in observed summer ice loss in the Arctic through 2016.

The choice of CMIP6 ice edge analysis technique as a model bias diagnostic was shown to be mostly arbitrary between the SPS and the mean IIEE. The exception to this is the case of single member simulations. Because not all CMIP6 models provide multiple ensemble realizations, the deterministic models may appear less skillful by the SPS in representing the ice edge than their ensemble model counterparts (see GFDL-ESM4; Figs. 7a,b). While the benefits of using ensemble models are beyond argument, our application of the mean IIEE identified the same relative model biases as the SPS, albeit with larger magnitudes. As such, the IIEE approach may provide a more equitable comparison of an individual model's skill in randomly representing the sea ice edge in any single realization (Fig. S5).

RASM-G is examined alongside the CMIP6 models and achieves favorable results in all categories, so it is reasonable to consider it as a realistic alternative sea ice reference simulation of the multidecadal evolution of the Arctic sea ice, especially with regard to SIT and SIV. The RASM-G configuration is similar to PIOMAS, in which both models use atmospheric reanalysis to force the ocean and sea ice model components. However, PIOMAS assimilates SIC and sea surface temperature data in order to improve model performance of sea ice characteristics. Yet, assimilated SIC fields come with passive microwave uncertainties (Yang et al. 2016), described earlier, which might contribute to the reported PIOMAS SIT and SIV biases and uncertainty in its trend estimates (Lindsay 2010; Schweiger et al. 2011). RASM-G does not use data assimilation for its components and still performs remarkably well in replicating observed SIE, as well as comparable to PIOMAS SIT distribution and SIV time series over multiple decades. We attribute this in part to the realistic atmospheric forcing, as well

as to a more realistic representation of sea ice relevant processes (e.g., oceanic forcing along marginal ice zones) and high spatial and temporal resolution. This lends confidence in the RASM-G model physics yielding correct depiction of the mean state and evolution of the Arctic ice pack.

*Acknowledgments.* We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. We acknowledge partial support from the following programs: U.S. Navy (to MW), the Department of Energy (DOE) Regional and Global Model Analysis (RGMA) (DOE: 89243019SSC0036 and DESC0014117), the Office of Naval Research (ONR) Arctic and Global Prediction (AGP) (ONR: N0001418WX00364) and National Science Foundation (NSF) Arctic System Science (ARCSS) (to WM, YL, JCK) (NSF: IAA1417888 and IAA1603602), and Ministry of Science and Higher Education in Poland (to RO). The Department of Defense (DOD) High Performance Computer Modernization Program (HPCMP) provided computer resources for RASM-G simulations analyzed here.

*Data availability statement.* CMIP6, *CryoSat-2*, EUMETSAT, *ICESat*, NSIDC, and PIOMAS data used for this study can be acquired from the links provided in sections 2a and 2b. The RASM-G data can be acquired from the Naval Postgraduate School (<https://nps.app.box.com/folder/139647168752?s=xyp563ee40w6lffn4718zr52a6m7rtcc>).

## REFERENCES

- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78** (1), 1–3, [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2).
- Cai, L., V. A. Alexeev, and J. E. Walsh, 2020: Arctic sea ice growth in response to synoptic- and large-scale atmospheric forcing from CMIP5 models. *J. Climate*, **33**, 6083–6099, <https://doi.org/10.1175/JCLI-D-19-0326.1>.
- Cassano, J. J., and Coauthors, 2017: Development of the Regional Arctic System Model (RASM): Near-surface atmospheric climate sensitivity. *J. Climate*, **30**, 5729–5753, <https://doi.org/10.1175/JCLI-D-15-0775.1>.
- Comiso, J. C., W. N. Meier, and R. Gersten, 2017: Variability and trends in the Arctic sea ice cover: Results from different techniques. *J. Geophys. Res. Oceans*, **122**, 6883–6900, <https://doi.org/10.1002/2017JC012768>.
- Davy, R., and S. Outten, 2020: The Arctic surface climate in CMIP6: Status and developments since CMIP5. *J. Climate*, **33**, 8047–8068, <https://doi.org/10.1175/JCLI-D-19-0990.1>.
- DeRepentigny, P., L. B. Tremblay, R. Newton, and S. Pfirman, 2016: Patterns of sea ice retreat in the transition to a seasonally ice-free Arctic. *J. Climate*, **29**, 6993–7008, <https://doi.org/10.1175/JCLI-D-15-0733.1>.
- Eisenman, I., W. N. Meier, and J. R. Norris, 2014: A spurious jump in the satellite record: Has Antarctic sea ice expansion been overestimated? *Cryosphere*, **8**, 1289–1296, <https://doi.org/10.5194/tc-8-1289-2014>.

- 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>.
- 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>.
- Goessling, H. F., and T. Jung, 2018: A probabilistic verification score for contours: Methodology and application to Arctic ice-edge forecasts. *Quart. J. Roy. Meteor. Soc.*, **144**, 735–743, <https://doi.org/10.1002/qj.3242>.
- , S. Tietsche, J. J. Day, E. Hawkins, and T. Jung, 2016: Predictability of the Arctic sea ice edge. *Geophys. Res. Lett.*, **43**, 1642–1650, <https://doi.org/10.1002/2015GL067232>.
- Hamman, J., and Coauthors, 2016: Land surface climate in the Regional Arctic System Model. *J. Climate*, **29**, 6543–6562, <https://doi.org/10.1175/JCLI-D-15-0415.1>.
- Hendricks, S., and R. Ricker, 2019: Product user guide & algorithm specification: AWI CryoSat-2 sea ice thickness, version 2.2. EPIC3, accessed 18 September 2020, <https://epic.awi.de/id/eprint/50033>.
- IPCC, 2019: Special report: The ocean and cryosphere in a changing climate. Accessed 14 June 2020, <https://www.ipcc.ch/srocc/cite-report/>.
- Ivanova, N., O. M. Johannessen, L. T. Pedersen, and R. T. Tonboe, 2014: Retrieval of Arctic sea ice parameters by satellite passive microwave sensors: A comparison of eleven sea ice concentration algorithms. *IEEE Trans. Geosci. Remote Sens.*, **52**, 7233–7246, <https://doi.org/10.1109/TGRS.2014.2310136>.
- Jackson, J. M., S. E. Allen, F. A. McLaughlin, R. A. Woodgate, and E. C. Carmack, 2011: Changes to the near-surface waters in the Canada Basin, Arctic Ocean from 1993–2009: A basin in transition. *J. Geophys. Res. Oceans*, **116**, C10008, <https://doi.org/10.1029/2011JC007069>.
- Jahn, A., J. E. Kay, M. M. Holland, and D. M. Hall, 2016: How predictable is the timing of a summer ice-free Arctic? *Geophys. Res. Lett.*, **43**, 9113–9120, <https://doi.org/10.1002/2016GL070067>.
- Kwok, R., 2018: Arctic sea ice thickness, volume, and multiyear ice coverage: Losses and coupled variability (1958–2018). *Environ. Res. Lett.*, **13**, 105005, <https://doi.org/10.1088/1748-9326/aae3ec>.
- , and D. A. Rothrock, 2009: Decline in Arctic sea ice thickness from submarine and ICESat records: 1958–2008. *Geophys. Res. Lett.*, **36**, L15501, <https://doi.org/10.1029/2009GL039035>.
- Labe, Z., G. Magnusdottir, and H. Stern, 2018: Variability of Arctic sea ice thickness using PIOMAS and the CESM Large Ensemble. *J. Climate*, **31**, 3233–3247, <https://doi.org/10.1175/JCLI-D-17-0436.1>.
- Lavergne, T., and Coauthors, 2019: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records. *Cryosphere*, **13**, 49–78, <https://doi.org/10.5194/tc-13-49-2019>.
- Lind, S., R. B. Ingvaldsen, and T. Furevik, 2018: Arctic warming hotspot in the northern Barents Sea linked to declining sea-ice import. *Nat. Climate Change*, **8**, 634–639, <https://doi.org/10.1038/s41558-018-0205-y>.
- Lindsay, R., 2010: New unified sea ice thickness climate data record. *Eos Trans. Amer. Geophys. Union*, **91**, 405–406, <https://doi.org/10.1029/2010EO440001>.
- Maslowski, W., J. Clement Kinney, M. Higgins, and A. Roberts, 2012: The future of Arctic sea ice. *Annu. Rev. Earth Planet. Sci.*, **40**, 625–654, <https://doi.org/10.1146/annurev-earth-042711-105345>.
- , —, S. R. Okkonen, R. Osinski, A. F. Roberts, and W. J. Williams, 2014: The large scale ocean circulation and physical processes controlling Pacific–Arctic interactions. *The Pacific Arctic Region: Ecosystem Status and Trends in a Rapidly Changing Environment*, J. M. Grebeiner and W. Maslowski, Eds., Springer, 101–132.
- Massonnet, F., T. Fichefet, H. Goosse, C. M. Bitz, G. Philippon-Berthier, M. M. Holland, and P.-Y. Barriat, 2012: Constraining projections of summer Arctic sea ice. *Cryosphere*, **6**, 1383–1394, <https://doi.org/10.5194/tc-6-1383-2012>.
- Meier, W. N., and J. S. Stewart, 2019: Assessing uncertainties in sea ice extent climate indicators. *Environ. Res. Lett.*, **14**, 035005, <https://doi.org/10.1088/1748-9326/aaf52c>.
- , J. Stroeve, and F. Fetterer, 2007: Whither Arctic sea ice? A clear signal of decline regionally, seasonally and extending beyond the satellite record. *Ann. Glaciol.*, **46**, 428–434, <https://doi.org/10.3189/172756407782871170>.
- , G. Peng, D. J. Scott, and M. H. Savoie, 2014: Verification of a new NOAA/NSIDC passive microwave sea-ice concentration climate record. *Polar Res.*, **33**, 21 004, <https://doi.org/10.3402/polar.v33.21004>.
- , 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, <https://doi.org/10.7265/N59P2ZTG..>
- Melling, H., 2002: Sea ice of the northern Canadian Arctic Archipelago. *J. Geophys. Res.*, **107**, 3181, <https://doi.org/10.1029/2001JC001102>.
- 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., Math. Phys. Eng. Sci.*, **373**, 20140164, <https://doi.org/10.1098/rsta.2014.0164>.
- , and SIMIP Community, 2020: Arctic sea ice in CMIP6. *Geophys. Res. Lett.*, **47**, e2019GL086749, <https://doi.org/10.1029/2019GL086749>.
- , A. Jahn, M. Holland, E. Hunke, F. Massonnet, J. Stroeve, B. Tremblay, and M. Vancoppenolle, 2016: The CMIP6 Sea-Ice Model Intercomparison Project (SIMIP): Understanding sea ice through climate-model simulations. *Geosci. Model Dev.*, **9**, 3427–3446, <https://doi.org/10.5194/gmd-9-3427-2016>.
- Onarheim, I. H., T. Eldevik, L. H. Smedsrud, and J. C. Stroeve, 2018: Seasonal and regional manifestation of Arctic sea ice loss. *J. Climate*, **31**, 4917–4932, <https://doi.org/10.1175/JCLI-D-17-0427.1>.
- Parkinson, C., J. Comiso, H. Zwally, D. Cavalieri, P. Gloersen, and W. Campbell, 1987: Arctic sea ice, 1973–1976: Satellite passive-microwave observations. NASA, 206 pp.
- Proshutinsky, A., and Coauthors, 2019: Analysis of the Beaufort Gyre freshwater content in 2003–2018. *J. Geophys. Res. Oceans*, **124**, 9658–9689, <https://doi.org/10.1029/2019JC015281>.
- 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. Oceans*, **116**, C00D07, <https://doi.org/10.1029/2011JC007110>.
- Randall, D. A., and Coauthors, 2007: Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 589–662.

- Roberts, A., and Coauthors, 2015: Simulating transient ice-ocean Ekman transport in the Regional Arctic System Model and Community Earth System Model. *Ann. Glaciol.*, **56**, 211–228, <https://doi.org/10.3189/2015AoG69A760>.
- 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>.
- Schlichtholz, P., 2019: Subsurface ocean flywheel of coupled climate variability in the Barents Sea hotspot of global warming. *Sci. Rep.*, **9**, 13692, <https://doi.org/10.1038/s41598-019-49965-6>.
- Schweiger, A., R. Lindsay, J. Zhang, M. Steele, H. Stern, and R. Kwok, 2011: Uncertainty in modeled Arctic sea ice volume. *J. Geophys. Res. Oceans*, **116**, C00D06, <https://doi.org/10.1029/2011JC007084>.
- Screen, J. A., 2011: Sudden increase in Antarctic sea ice: Fact or artifact? *Geophys. Res. Lett.*, **38**, L13702, <https://doi.org/10.1029/2011GL047553>.
- 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>.
- , 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>.
- , A. P. Barrett, J. C. Stroeve, D. N. Kindig, and M. M. Holland, 2009: The emergence of surface-based Arctic amplification. *Cryosphere*, **3**, 11–19, <https://doi.org/10.5194/tc-3-11-2009>.
- Shen, Z., A. Duan, D. Li, and J. Li, 2021: Assessment and ranking of climate models in Arctic sea-ice cover simulation: From CMIP5 to CMIP6. *J. Climate*, **34**, 3609–3627, <https://doi.org/10.1175/JCLI-D-20-0294.1>.
- Shu, Q., Q. Wang, Z. Song, F. Qiao, J. Zhao, M. Chu, and X. Li, 2020: Assessment of sea ice extent in CMIP6 with comparison to observations and CMIP5. *Geophys. Res. Lett.*, **47**, e2020GL087965, <https://doi.org/10.1029/2020GL087965>.
- Stroeve, J., 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>.
- , and —, 2018: Changing state of Arctic sea ice across all seasons. *Environ. Res. Lett.*, **13**, 103001, <https://doi.org/10.1088/1748-9326/aade56>.
- , M. C. Serreze, M. M. Holland, J. E. Kay, J. Malanik, and A. P. Barrett, 2012: 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>.
- , A. Barrett, M. Serreze, and A. Schweiger, 2014: Using records from submarine, aircraft and satellites to evaluate climate model simulations of Arctic sea ice thickness. *Cryosphere*, **8**, 1839–1854, <https://doi.org/10.5194/tc-8-1839-2014>.
- Taylor, P. C., M. Cai, A. Hu, J. Meehl, W. Washington, and G. J. Zhang, 2013: A decomposition of feedback contributions to polar warming amplification. *J. Climate*, **26**, 7023–7043, <https://doi.org/10.1175/JCLI-D-12-00696.1>.
- Timmermans, M.-L., J. Toole, and R. Krishfield, 2018: Warming of the interior Arctic Ocean linked to sea ice losses at the basin margins. *Sci. Adv.*, **4**, eaat6773, <https://doi.org/10.1126/sciadv.aat6773>.
- Tsujino, H., and Coauthors, 2018: JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do). *Ocean Modell.*, **130**, 79–139, <https://doi.org/10.1016/j.ocemod.2018.07.002>.
- Wayand, N. E., C. M. Bitz, and E. Blanchard-Wrigglesworth, 2019: A year-round subseasonal-to-seasonal sea ice prediction portal. *Geophys. Res. Lett.*, **46**, 3298–3307, <https://doi.org/10.1029/2018GL081565>.
- 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>.
- Yang, Q., M. Losch, S. N. Losa, T. Jung, L. Nerger, and T. Lavergne, 2016: Brief communication: The challenge and benefit of using sea ice concentration satellite data products with uncertainty estimates in summer sea ice data assimilation. *Cryosphere*, **10**, 761–774, <https://doi.org/10.5194/tc-10-761-2016>.
- Yi, D., and H. J. Zwally, 2009: Arctic sea ice freeboard and thickness, version 1, National Snow and Ice Data Center Distributed Active Archive Center, accessed 28 August 2020, <https://doi.org/10.5067/SXJVJ3A2XIZT>.
- Zhang, J., and D. A. Rothrock, 2003: Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Mon. Wea. Rev.*, **131**, 845–861, [https://doi.org/10.1175/1520-0493\(2003\)131<0845:MGSIWA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<0845:MGSIWA>2.0.CO;2).
- Zygmuntowska, M., P. Rampal, N. Ivanova, and L. H. Smedsrud, 2013: Uncertainties in Arctic sea ice thickness and volume: New estimates and implications for trends. *Cryosphere Discuss.*, **7**, 5051–5095, <https://doi.org/10.5194/tcd-7-5051-2013>.