Assessment of the Pan-Arctic Accelerated Rate of Sea Ice Decline in CMIP6 Historical Simulations

YOUNJOO J. LEE,a MATTHEW WATTS,a WIESLAW MASLOWSKI,a JACLYN CLEMENT KINNEY,a AND ROBERT OSINSKIb

a Department of Oceanography, Naval Postgraduate School, Monterey, California
b Institute of Oceanology, Polish Academy of Sciences, Sopot, Poland

Abstract: Arctic sea ice loss in response to a warming climate is assessed in 42 models participating in phase 6 of the Coupled Model Intercomparison Project (CMIP6). Sea ice observations show a significant acceleration in the rate of decline commencing near the turn of the twenty-first century. It is our assertion that state-of-the-art climate models should qualitatively reflect this accelerated trend within the limitations of internal variability and observational uncertainty. Our analysis shows that individual CMIP6 simulations of sea ice depict a wide range of model spread on biases and anomaly trends both across models and among their ensemble members. While the CMIP6 multimodel mean captures the observed sea ice area (SIA) decline relatively well, an individual model’s ability to represent the acceleration in sea ice decline remains a challenge. Seventeen (40%) out of 42 CMIP6 models and 37 (13%) out of the total 286 ensemble members reasonably capture the observed trend and acceleration in SIA decline. In addition, a larger ensemble size appears to increase the odds for a model to include at least one ensemble member skillfully representing the accelerated SIA trends. Simulations of sea ice volume (SIV) show much larger spread and uncertainty than SIA; however, due to limited availability of sea ice thickness data, these are not as well constrained by observations. Finally, we find that models with more ocean heat transport simulate larger sea ice declines, which suggests an emergent constraint in CMIP6 ensembles. This relationship points to the need for better understanding and modeling of ice–ocean interactions, especially with respect to frazil ice growth.

Keywords: Arctic; Sea ice; Climate models; Model comparison; Ocean; Climate variability

1. Introduction

The impact of climate change has been most evident in high latitudes where accelerated sea ice loss and warming of surface air temperatures have been documented (Serreze et al. 2009; Taylor et al. 2013; IPCC 2019). Relative to the global average, the Arctic surface climate exhibits an amplified response to forced greenhouse gas warming, known as Arctic amplification (AA). It has primarily been attributed to several positive feedbacks within the atmosphere–ocean–sea ice system (e.g., Previdi et al. 2021). The effects of AA can be seen in the negative trend in sea ice cover (Serreze and Francis 2006; Dai et al. 2019) observed through all months of the satellite record since 1978 (Serreze and Barry 2011; Stroeve and Notz 2018). On the other hand, the continuing summer decline of sea ice reduces surface albedo to increase the absorption of solar radiation and warming in the upper ocean, which delays sea ice formation and warms the atmosphere in fall and winter, thus contributing to AA. This is partly why the leading causes of amplified warming in the Arctic are still under debate (Pithan and Mauritsen 2014; Stuecker et al. 2018).

Under a reduced sea ice cover regime, the Arctic Ocean not only absorbs more shortwave radiation but also exchanges additional momentum and heat across the ocean–atmosphere interface (Parkinson et al. 1987; Rampal et al. 2011; Proshutinsky et al. 2019). Hence, alterations in surface energy fluxes due to climate change have significant impacts on boundary layer processes such as ice–albedo feedback (e.g., Hudson et al. 2013) and Arctic cloud formation (e.g., Walsh et al. 2009). As sea ice plays a crucial role in Earth’s energy budget (Hartmann and Ceppi 2014; Pistone et al. 2014; Riihelä et al. 2021), it is essential for Earth system and global climate models (hereafter climate models) to accurately simulate the changing sea ice state together with shifting climate.

Simulations of historical Arctic sea ice over the past several phases of the Coupled Model Intercomparison Project (CMIP) have shown similar performance (e.g., spread in mean values, seasonal cycles, and long-term trends) with respect to the large-scale integrated measures of sea ice area (SIA), extent (SIE), and volume (SIV) (Stroeve et al. 2007, 2012a; Notz et al. 2020; Shu et al. 2020; Davy and Outten 2020; Shen et al. 2021). While large model spread in these integrated measures endures, the CMIP multimodel mean (MMM) continues to outperform all individual models (e.g., Watts et al. 2021). Some progress has been reported in phase 6 of CMIP (CMIP6) MMM compared to previous CMIP activity, including the observed sensitivity of Arctic sea ice to anthropogenic CO2 emissions and global warming (Notz et al. 2020), with SIE trends closer to observations (Davy and Outten 2020; Shen et al. 2021). However, much uncertainty remains as to whether this progress in CMIP6 can be attributed to upgraded model physics, changes in external forcing (cf. DeRepentigny et al. 2022), or “by chance.”
variability. Still, CMIP6 simulations are seemingly unable to capture the faster decline in September sea ice observed for 1979–2014 (Davy and Outten 2020; Shen et al. 2021), especially after 2000 (Shu et al. 2020), and global mean surface temperature simultaneously (Notz et al. 2020).

In this contribution, we assess the decline of sea ice in CMIP6 historical simulations (1979–2014) and compare model results to the observed sea ice trend, which is a manifestation of coupled atmosphere–sea ice interactions over the Arctic (e.g., Matsumura et al. 2014; Huang et al. 2019; Liu et al. 2021). The observed sea ice decline is characterized by a gradual loss of pan-Arctic sea ice cover prior to the late 1990s, followed by an enhanced rate of decline through the 2000s (Comiso et al. 2008; Serreze and Stroeve 2015). Although the causes are not well recognized, it is suggested that biomass burning emissions can influence simulated Arctic sea ice trends through aerosol–cloud interactions (DeRepentigny et al. 2022). Thereafter, recent studies have shown a slowdown of sea ice decline (Baxter et al. 2019; Perovich et al. 2020; Zhang 2021). This motivates us to examine how well this accelerated sea ice decline is represented in CMIP6 models during the historical period in terms of timing and magnitude and, if so, what potential causes are linked to it. However, one must be cautious when evaluating models on shorter time scales (i.e., years), especially during rapid or slow ice decline, because internal climate variability plays a role in determining model spread (Swart et al. 2015).

In this study, we focus on the acceleration of negative sea ice trends in recent decades by examining the entire time series in Arctic SIA and SIV and identifying statistical significance in trends among CMIP6 models instead of seasonal sea ice maximum and minimum (March and September, respectively) time series. Although previous studies (e.g., Notz et al. 2020) have provided insights on sea ice trends in the latest CMIP models, we present a different perspective on sea ice trends and acceleration by minimizing the apparent overlap. Furthermore, a new model evaluation method (i.e., emergent constraint approach; Hall et al. 2019) is introduced and discussed in relation to ocean–ice feedback to offer insights into the potential reduction of uncertainties in CMIP6 sea ice simulations regarding recent accelerated sea ice decline and ocean heat transport.

2. Data and methods

a. Model output

Our study uses CMIP6 historical experiment data to evaluate Arctic sea ice during the recent past in climate models (Notz et al. 2016; Eyring et al. 2016). We retrieved sea ice outputs from 42 CMIP6 models, including all 286 of their ensemble members, from the Earth System Grid Federation (ESGF) repository (https://esgf-node.llnl.gov/search/cmip6/). Additionally, depth-integrated net northward ocean heat transport (OHT) output (a field defined as hbasin) only available for 17 out of 42 CMIP6 models was retrieved from ESGF (Griffies et al. 2016). Table S1 in the online supplemental material summarizes the model simulations and variables used for this analysis.

Due to the lack of persistent SIT observations over the Arctic, a long-term observational time series of SIV does not exist. Instead, we use the CMIP6 SIV MMM as the primary reference for model evaluation. Additionally, two SIV reanalyses are included for further comparison against CMIP6 models: the Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS; Schweiger et al. 2011) and an ice–ocean version of the Regional Arctic System Model (RASM-G; Maslowski et al. 2012; Roberts et al. 2015; Hamman et al. 2016; Cassano et al. 2017) forced with the Japanese 55-year Reanalysis (Kobayashi et al. 2015). The monthly mean PIOMAS SIV reanalysis (version 2.1) for 1979–2014 was retrieved from the Polar Science Center at the University of Washington (http://psc.apl.uw.edu/research/projects/arctic-sea-ice-volume-anomaly/data). RASM-G is used as an alternative SIV model reanalysis available at the Naval Postgraduate School.

b. Observational data

To compare model simulations of sea ice against observations, we retrieved monthly mean SIC data for the period 1979–2014 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). Following Notz et al. (2020), Olason and Notz (2014), and the IPCC’s Sixth Assessment Report (2021), a mean observational reference SIC was determined by combining the National Aeronautics and Space Administration (NASA) Team (NT) and Bootstrap (BT) SIC algorithms [Cavalieri et al. (1984) and Comiso (1986), respectively], NOAA/NSIDC climate data record (CDR) of Passive Microwave SIC, version 4 (Meier et al. 2021), and the OSI SAF team SIC algorithm (OSI-450; Lavergne et al. 2019; OSI SAF 2017). The OSI SAF SIC data were first linearly interpolated onto the NSIDC grid and then averaged with other algorithm data to create our primary observational reference. Monthly mean SIA is calculated by multiplying the gridcell area by SIC. The spread in observational estimates (i.e., standard deviation) due to algorithm differences can be interpreted as observational uncertainty (Meier and Stewart 2019; Notz et al. 2020). Unless specified otherwise, the observed sea ice trend is referred to as the averaged SIA from the four different SIC algorithms.

c. Methods

Time series linear trend analysis is used to examine the monthly mean pan-Arctic SIA and SIV for the period of 1979–2014 and two selected subperiods, P1 (1979–96) and P2 (1997–2014). The sea ice loss is generally best described by linear regression models for at least the last 30 years (Peng et al. 2020). Here, the simulated SIA time series for CMIP6 models was calculated as the product of SIC (sicone) and grid-cell area (areacell) for all ocean grid cells in the Northern Hemisphere. If these variables were not provided, we used the variable called Northern Hemisphere SIA (siareal). Analogously, we computed simulated SIV time series by taking the product of SIV (sivol) or sea ice mass (simass) divided by the density of sea ice ($\rho_{si}$) and areacell for all Northern
Hemisphere ocean grid cells. If not provided, we used the Northern Hemisphere SIV (sivolu) instead. Following Notz et al. (2020), SIA was analyzed as our 2D sea ice evaluation metric over SIE (i.e., sum of gridcell areas with SIC greater than 15%), which is a strong grid-dependent metric (Notz 2014). Nevertheless, we also calculated SIE and determined the average difference in sea ice trends between the two metrics to be less than 5% for CMIP6 MMM (individual model differences range from 0.1% to 24%; Table S2).

Characteristic anomalies are all calculated relative to the 1979–2014 monthly mean for individual models and observational references by removing their mean annual cycles. Assessing monthly anomalies allows us to remove the seasonal signal that could otherwise distort the statistics and reduce the influence of individual model biases when examining trend behaviors of the simulated ice pack. Since the majority of CMIP6 models provided multiple ensemble members (up to 32; Table S1), all results are based on model ensemble means unless otherwise noted. Also, a CMIP6 MMM was calculated for SIA, SIV, and OHT; we averaged the individual model ensemble means instead of all ensemble members to avoid biasing the MMM toward the large ensemble models.

Model and observed trends (βu and βo, respectively) are determined by least squares linear regression of the monthly mean anomaly time series. Trend uncertainties were calculated following Santer et al. (2008) and Stroeve et al. (2012a), whereby we adjusted the modeled and observed standard errors [s(βu) and s(βo), respectively] using an effective sample size (n_eff) to account for the large lag-1 temporal autocorrelation (AR1) of the trend residuals:

\[ n_{\text{eff}} = n_{\text{tot}}(1 - \text{AR1})/(1 + \text{AR1}), \]

where n_tot is the number of total months over which the trend is calculated. This is necessary because many geophysical data show pronounced month-to-month persistence, and as such are not statistically independent. Essentially, larger AR1 reduces the number of statistically independent samples, therefore decreasing the statistical degrees of freedom and increasing trend uncertainty.

Furthermore, assuming a normal distribution (Santer et al. 2008; Stroeve et al. 2012a), we applied a paired two-tailed t test against p = 0.10, in which the observed trend (βo) is tested against each model realization trend (βm) to reject the null hypothesis, which states that the model trend is no different than the observed trend. Here, d is the normalized difference between the trends in any two modeled and observed time series, and s(βm) and s(βo) are the standard errors of βm and βo, respectively:

\[ d = (\beta_m - \beta_o)/\sqrt{s(\beta_m)^2 + s(\beta_o)^2}. \]

3. Results

Similar to previous studies (Comiso et al. 2008; Stroeve et al. 2012b; Serreze and Stroeve 2015), we detect an accelerated decline of SIA in the combined passive microwave observations (Fig. 1a). The P2 rate of SIA decline in the observations [−0.69 × 10^6 km² decade⁻¹] with an adjusted 95% confidence interval (CI) of [±0.26] represents a 72% increase compared to the P1 rate (−0.39 × 10^6 km² decade⁻¹) with an adjusted 95% CI of [±0.23]. Our analysis also shows an increased rate of SIV decline in the PIOMAS reanalysis (Fig. 1b). Here, the P2 rate (−4.74 × 10^3 km³ decade⁻¹) with an adjusted 95% CI of [±1.40] represents a 218% increase compared to the P1 rate (−1.49 × 10^3 km³ decade⁻¹) with an adjusted 95% CI of [±2.34].

a. 1979–2014 mean state and trends of simulated Arctic sea ice

The CMIP6 MMM SIA consistently overestimates the observed SIA with a mean bias of 0.71 × 10^6 km² (Fig. 2a; see also Fig. S1a for individual model SIA time series) as the majority of the models (29 out of 42) overestimate it. For the majority of CMIP6 models, mean SIA ranges between 9 and 14 × 10^6 km² in P1 and 8 and 13 × 10^6 km² in P2 (Fig. 2a), and their time series remain within one standard deviation (σ) of the CMIP6 MMM. However, six CMIP6 models (GISS-E2.1H, BCC-CSM 2 MR, CAMS-CSM 1.0, E3SM 1.0, and CIESM) are notably biased, showing time series outside of the ±1σ range from the CMIP6 MMM for the total period (Fig. 2a). Furthermore, GISS-E2.1H and BCC-CSM 2 MR show the largest positive bias, and CIESM exhibits the most prominent negative bias.

All models simulate a decline in SIA with varying intensity, in general agreement with the observed historical decline for 36 years (Fig. 2c and Table 1). However, the majority of CMIP6 models (27 out of 42) underestimate the SIA trend, and most of those models (14 out of 15) have a positive SIA bias (Figs. 2b,c). The individual model trends exhibit a large spread, ranging from less than one quarter (e.g., NorESM2-MM at −0.13 × 10^6 km² decade⁻¹) to approximately twice (e.g., E3SM 1.0 at −1.01 × 10^6 km² decade⁻¹) the observed trend. Despite large intermodel variability, the CMIP6 SIA MMM rate of decline (−0.47 × 10^6 km² decade⁻¹) underestimates the observed trend (−0.54 × 10^6 km² decade⁻¹) by 0.07 × 10^6 km² decade⁻¹. Individual ensemble member trends are shown in Fig. S2a.

Similar to the CMIP5 evaluation (see Stroeve et al. 2014), all models simulate a decline in SIV with varying intensity, which qualitatively matches the reanalyses and satellite-estimated (Kwok 2018) SIV trends, but the rates seem to be faster in models with thicker ice than in those with thinner ice (Fig. 3a). Note that one model, NorCPM1, is excluded from the CMIP6 SIV MMM because it has the highest bias (40 × 10^3 km³; more than twice as large as any other model; Fig. 3b). For 30 out of 42 CMIP6 models, SIV ranges within ±1σ from the MMM (i.e., 11−35 × 10^3 km³) but models exhibit much stronger intermodel spread across a broader range in SIV (Fig. 3a; at least a factor of 7.2–9.1 between the highest and lowest estimates), compared to SIA (Fig. 2a; a factor of 2.2–2.5 between the highest and lowest estimates). Beyond around the year 2000, the model spread in simulated SIV is reduced by 45% to a minimum range of 11–22 × 10^3 km³ (Fig. 3a). This indicates that larger simulated SIT uncertainty exists during the early portion
of our analysis period. Models with the largest bias are E3SM 1.1 ECA, the largest positively biased, and CIESM, the largest negatively biased aside from NorCPM1 (Fig. 3b; see Fig. S1 for individual model SIV time series).

Collectively, the CMIP6 SIV MMM mean of $21.4 \times 10^3 \text{ km}^3$ and trend of $2.88 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$ for 1979–2014 pair well with the two SIV reanalyses (Figs. 3b,c; Table 2). Relative to the CMIP6 SIV MMM mean, PIOMAS has a slight negative difference in mean ($20.2 \times 10^3 \text{ km}^3$) and trend ($-3.03 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$), while RASM-G has no difference in mean and a slower trend ($-2.52 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$). Individual CMIP6 SIV trends range from $-0.68 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$ (CNRM-CM6-1) to $-5.66 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$ (E3SM 1.0). Individual ensemble member trends are shown in Fig. S2b.

b. Sea ice variability

To evaluate CMIP6 model skill in simulating sea ice variability, Taylor diagrams (Taylor 2001) are used to quantify the
statistics of CMIP6 sea ice anomalies against specific references: the combined satellite observed SIA and CMIP6 SIV MMM (Fig. 4). This analysis essentially removes an individual model bias and presents the following statistics: the correlation coefficient ($r$) measures the strength of linear relationship between simulated and reference values, the unbiased root-mean-square difference (uRMSD) describes the difference between CMIP models and references, and the normalized standard deviation indicates the ratio of model variability against the reference. Additionally, individual SIA observational estimates (i.e., CDR,
<table>
<thead>
<tr>
<th>Model</th>
<th>Full (Model)</th>
<th>Mean</th>
<th>Adjusted CI (P1)</th>
<th>Mean</th>
<th>Adjusted CI (P2)</th>
<th>Mean</th>
<th>Adjusted CI (P2/P1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCESS-CM2</td>
<td>(0/3)</td>
<td>-0.33</td>
<td>0.12</td>
<td>-0.22</td>
<td>0.29</td>
<td>-0.23</td>
<td>0.32</td>
</tr>
<tr>
<td>ACCESS-ESM1.5</td>
<td>(2/20)</td>
<td>-0.37</td>
<td>0.12</td>
<td>-0.20</td>
<td>0.30</td>
<td>-0.33</td>
<td>0.31</td>
</tr>
<tr>
<td>AWI-CM 1.1 MR</td>
<td>(0/5)</td>
<td>-0.37</td>
<td>0.12</td>
<td>-0.18</td>
<td>0.29</td>
<td>-0.46</td>
<td>0.26</td>
</tr>
<tr>
<td>AWI-ESM 1.1 LR</td>
<td>(0/1)</td>
<td>-0.32</td>
<td>0.08</td>
<td>-0.46</td>
<td>0.22</td>
<td>-0.38</td>
<td>0.22</td>
</tr>
<tr>
<td>BCC-CSM 2.1 MR</td>
<td>(1/3)</td>
<td>-0.52</td>
<td>0.17</td>
<td>-0.29</td>
<td>0.48</td>
<td>-0.63</td>
<td>0.44</td>
</tr>
<tr>
<td>BCC-ESM 1</td>
<td>(1/3)</td>
<td>-0.44</td>
<td>0.12</td>
<td>-0.31</td>
<td>0.36</td>
<td>-0.49</td>
<td>0.27</td>
</tr>
<tr>
<td>CAMS-CSM 1.0</td>
<td>(0/2)</td>
<td>-0.24</td>
<td>0.09</td>
<td>-0.34</td>
<td>0.23</td>
<td>-0.18</td>
<td>0.25</td>
</tr>
<tr>
<td>CanESM5</td>
<td>(7/25)</td>
<td>-0.82</td>
<td>0.12</td>
<td>-0.50</td>
<td>0.28</td>
<td>-1.04</td>
<td>0.26</td>
</tr>
<tr>
<td>CanESM5-CanOE</td>
<td>(1/3)</td>
<td>-0.78</td>
<td>0.18</td>
<td>-0.16</td>
<td>0.28</td>
<td>-1.27</td>
<td>0.23</td>
</tr>
<tr>
<td>CAS-ESM2.0</td>
<td>(0/4)</td>
<td>-0.41</td>
<td>0.09</td>
<td>-0.19</td>
<td>0.27</td>
<td>-0.51</td>
<td>0.21</td>
</tr>
<tr>
<td>CESM2</td>
<td>(0/11)</td>
<td>-0.52</td>
<td>0.16</td>
<td>-0.20</td>
<td>0.37</td>
<td>-0.47</td>
<td>0.47</td>
</tr>
<tr>
<td>CESM2-FV2</td>
<td>(0/3)</td>
<td>-0.49</td>
<td>0.14</td>
<td>-0.08</td>
<td>0.38</td>
<td>-0.52</td>
<td>0.32</td>
</tr>
<tr>
<td>CESM2-WACCM</td>
<td>(0/3)</td>
<td>-0.56</td>
<td>0.13</td>
<td>-0.34</td>
<td>0.32</td>
<td>-0.49</td>
<td>0.37</td>
</tr>
<tr>
<td>CESM2-WACCM-FV2</td>
<td>(0/3)</td>
<td>-0.59</td>
<td>0.13</td>
<td>-0.38</td>
<td>0.32</td>
<td>-0.57</td>
<td>0.40</td>
</tr>
<tr>
<td>CIESM</td>
<td>(0/3)</td>
<td>-0.23</td>
<td>0.11</td>
<td>-0.06</td>
<td>0.33</td>
<td>-0.17</td>
<td>0.25</td>
</tr>
<tr>
<td>CNRM-CM6-1</td>
<td>(1/20)</td>
<td>-0.24</td>
<td>0.11</td>
<td>-0.22</td>
<td>0.28</td>
<td>-0.21</td>
<td>0.29</td>
</tr>
<tr>
<td>CNRM-CM6-1-HR</td>
<td>(0/1)</td>
<td>-0.29</td>
<td>0.09</td>
<td>-0.01</td>
<td>0.23</td>
<td>-0.41</td>
<td>0.24</td>
</tr>
<tr>
<td>CNRM-ESM2-1</td>
<td>(0/5)</td>
<td>-0.35</td>
<td>0.12</td>
<td>-0.25</td>
<td>0.36</td>
<td>-0.38</td>
<td>0.34</td>
</tr>
<tr>
<td>ESM1.0</td>
<td>(0/5)</td>
<td>-1.01</td>
<td>0.25</td>
<td>-0.25</td>
<td>0.42</td>
<td>-1.41</td>
<td>0.49</td>
</tr>
<tr>
<td>ESM1.1</td>
<td>(0/1)</td>
<td>-0.69</td>
<td>0.16</td>
<td>0.04</td>
<td>0.33</td>
<td>-0.84</td>
<td>0.24</td>
</tr>
<tr>
<td>ESM1.1 ECA</td>
<td>(0/1)</td>
<td>-0.66</td>
<td>0.18</td>
<td>0.19</td>
<td>0.26</td>
<td>-0.83</td>
<td>0.35</td>
</tr>
<tr>
<td>EC-Earth3</td>
<td>(0/9)</td>
<td>-0.65</td>
<td>0.14</td>
<td>-0.40</td>
<td>0.35</td>
<td>-0.83</td>
<td>0.33</td>
</tr>
<tr>
<td>EC-Earth3-Veg</td>
<td>(2/7)</td>
<td>-0.59</td>
<td>0.14</td>
<td>-0.39</td>
<td>0.29</td>
<td>-0.91</td>
<td>0.26</td>
</tr>
<tr>
<td>FGALS-ESM3-L</td>
<td>(0/3)</td>
<td>-0.36</td>
<td>0.12</td>
<td>-0.25</td>
<td>0.26</td>
<td>-0.33</td>
<td>0.31</td>
</tr>
<tr>
<td>GFDFL-CM4</td>
<td>(0/1)</td>
<td>-0.49</td>
<td>0.12</td>
<td>-0.52</td>
<td>0.33</td>
<td>-0.76</td>
<td>0.27</td>
</tr>
<tr>
<td>GFDFL-ESM4</td>
<td>(0/1)</td>
<td>-0.35</td>
<td>0.11</td>
<td>-0.30</td>
<td>0.45</td>
<td>-0.21</td>
<td>0.18</td>
</tr>
<tr>
<td>GISS-E21H</td>
<td>(2/10)</td>
<td>-0.74</td>
<td>0.18</td>
<td>-0.36</td>
<td>0.43</td>
<td>-0.84</td>
<td>0.44</td>
</tr>
<tr>
<td>HadGEM3-GC31-LL</td>
<td>(0/3)</td>
<td>-0.54</td>
<td>0.17</td>
<td>0.06</td>
<td>0.31</td>
<td>-0.92</td>
<td>0.24</td>
</tr>
<tr>
<td>HadGEM3-GC31-MM</td>
<td>(1/4)</td>
<td>-0.47</td>
<td>0.14</td>
<td>-0.24</td>
<td>0.27</td>
<td>-0.59</td>
<td>0.30</td>
</tr>
<tr>
<td>IPSL-CM6A-LR</td>
<td>(5/32)</td>
<td>-0.54</td>
<td>0.16</td>
<td>-0.31</td>
<td>0.39</td>
<td>-0.57</td>
<td>0.40</td>
</tr>
<tr>
<td>KIOST-ESM</td>
<td>(1/1)</td>
<td>-0.25</td>
<td>0.09</td>
<td>-0.24</td>
<td>0.22</td>
<td>-0.54</td>
<td>0.22</td>
</tr>
<tr>
<td>MIROC6</td>
<td>(1/10)</td>
<td>-0.32</td>
<td>0.09</td>
<td>-0.10</td>
<td>0.19</td>
<td>-0.44</td>
<td>0.24</td>
</tr>
<tr>
<td>MPI-ESM1.2-HAM</td>
<td>(0/3)</td>
<td>-0.38</td>
<td>0.15</td>
<td>0.00</td>
<td>0.40</td>
<td>-0.31</td>
<td>0.35</td>
</tr>
<tr>
<td>MPI-ESM1.2-RR</td>
<td>(3/10)</td>
<td>-0.41</td>
<td>0.14</td>
<td>-0.26</td>
<td>0.34</td>
<td>-0.51</td>
<td>0.34</td>
</tr>
<tr>
<td>MPI-ESM1.2-LR</td>
<td>(2/10)</td>
<td>-0.36</td>
<td>0.10</td>
<td>-0.24</td>
<td>0.27</td>
<td>-0.40</td>
<td>0.24</td>
</tr>
<tr>
<td>MRI-ESM2.0</td>
<td>(1/5)</td>
<td>-0.43</td>
<td>0.12</td>
<td>-0.29</td>
<td>0.28</td>
<td>-0.38</td>
<td>0.36</td>
</tr>
<tr>
<td>NorCPM1</td>
<td>(2/30)</td>
<td>-0.20</td>
<td>0.08</td>
<td>-0.08</td>
<td>0.18</td>
<td>-0.21</td>
<td>0.18</td>
</tr>
<tr>
<td>NorESM2-LM</td>
<td>(0/3)</td>
<td>-0.25</td>
<td>0.08</td>
<td>-0.27</td>
<td>0.20</td>
<td>-0.02</td>
<td>0.20</td>
</tr>
<tr>
<td>NorESM2-MM</td>
<td>(0/3)</td>
<td>-0.13</td>
<td>0.08</td>
<td>-0.16</td>
<td>0.19</td>
<td>-0.04</td>
<td>0.22</td>
</tr>
<tr>
<td>SAM6-UNICON</td>
<td>(0/1)</td>
<td>-0.42</td>
<td>0.14</td>
<td>-0.02</td>
<td>0.22</td>
<td>-0.90</td>
<td>0.31</td>
</tr>
<tr>
<td>TaiESM1</td>
<td>(0/1)</td>
<td>-0.61</td>
<td>0.15</td>
<td>-0.08</td>
<td>0.24</td>
<td>-0.97</td>
<td>0.41</td>
</tr>
<tr>
<td>UKESM1.0-LL</td>
<td>(4/14)</td>
<td>-0.59</td>
<td>0.12</td>
<td>-0.25</td>
<td>0.26</td>
<td>-0.72</td>
<td>0.24</td>
</tr>
<tr>
<td>CMIP6 SIA MMM</td>
<td>-0.46</td>
<td>0.06</td>
<td>-0.22</td>
<td>0.14</td>
<td>-0.55</td>
<td>0.10</td>
<td>2.55</td>
</tr>
<tr>
<td>Observations</td>
<td>-0.54</td>
<td>0.09</td>
<td>-0.39</td>
<td>0.23</td>
<td>-0.69</td>
<td>0.26</td>
<td>1.77</td>
</tr>
</tbody>
</table>

Eighty-one percent of the CMIP6 SIA ensemble simulations have anomaly correlation coefficients greater than 0.70 (the lower end of strong correlation, \( p < 0.05 \)), and 45% also have standard deviation while the NT and OSI-450 algorithms have a slight negative bias and larger standard deviation (not shown).

BT, NT, and OSI-450 algorithms are included to quantify uncertainties resulting from choice of algorithms. Compared to the combined SIA observations, all individual observational estimates show strong correlation coefficients (\( r = 0.99, p < 0.01 \)). The BT and CDR algorithms have a slight positive bias with a smaller
The majority of CMIP6 models (64%) show lower SIA variability (i.e., normalized standard deviation is less than one) than in observations, which is manifested in the CMIP6 SIA MMM (Fig. 4a). We also examine the detrended SIA anomalies (Fig. S3a) to evaluate model interannual variability without the influence of the negative long-term SIA trend. When detrended, all CMIP6 models shift left in the Taylor diagram, indicating a lower correlation. This suggests that the higher model skill identified in Fig. 4a is mostly controlled by larger normalized standard deviations between 0.75 and 1.25 (Fig. 4a).
covariance and lower standard deviation resulting from strong long-term decline of SIA. Only 12 out of 42 models exceed weak correlation ($r > 0.3$, $p < 0.05$) against the observed SIA anomaly and have normalized standard deviations between 0.75 and 1.25. Also, 10 out of these 12 models show statistically significant trends against the observed SIA trend (Fig. 2c).

When evaluating CMIP6 models against the CMIP6 SIV MMM anomaly for the period 1979–2014, all models, except CIESM and KIOST-ESM, are strongly correlated ($r > 0.70$, $p < 0.05$) with the CMIP6 MMM, but show larger spread in SIV variability compared to the SIA spread (Fig. 4b). When the same normalized standard deviation criterion used for SIA is applied between 0.75 and 1.25, we find 11 out of 42 CMIP6 models reasonably simulate the variability of the CMIP6 SIV MMM. However, when detrended, there is a large model spread and weaker correlation with the CMIP6 SIV MMM (Fig. 5b).
indicating stronger long-term decline in SIV. Also shown are the two SIV reanalyses, i.e., PIOMAS and RASM-G, which have a strong correlation and comparable standard deviation with the CMIP6 MMM.

c. Accelerated rates of sea ice decline

Following Santer et al. (2008) and Stroeve et al. (2012a), simulated SIA trends for all individual CMIP6 ensemble members are evaluated against observations for periods P1 and P2. There is a large range in SIA trends among models as well as within individual model’s ensemble members, the latter indicating sizeable internal variability (Fig. 5). The model spread is much larger when analyzing shorter climate periods (i.e., P1 and P2; Figs. 5a,b) rather than the full historical simulation period (Fig. S2a). Another consequence of examining shorter periods P1 and P2 is the sizable trend uncertainty (i.e., large error bars) that results from substantial lag-1 autocorrelation in the SIA time series (i.e., AR1 ranges from 0.52 to 0.89) and the subsequent adjustment of independent samples size (see section 2c).
To assess the CMIP6 models' skill in representing the multi-decadal sea ice variability and trends, a paired $t$ test is performed between each CMIP6 model ensemble member and the observed SIA trends (multialgorithm mean) separately for each subperiod. Although the P1 and P2 model trends fall within the range of the observed trends at an adjusted 95% CI, it is strongly recommended to carry out hypothesis testing to judge the significance of difference (Schenker and Gentleman 2001; Lanzante 2005). A paired $t$ test confirms that the observed sea ice decline is accelerated in the multialgorithm mean SIA anomaly, which is significant at $p = 0.0917$, with a ratio of P2/P1 trend $= 1.77$ (Table 1). Accounting for the uncertainty in the observed SIA, the individual accelerated trends (ratio of P2/P1) derived from the four different satellite algorithms are found to be between 1.44 and 2.09.

Out of the total 286 CMIP6 ensemble members, only 37 (13%) exhibit the observed trends and acceleration in SIA decline at the 90% CI ($p \leq 0.10$). Out of 42 CMIP6 models,
17 (40%) models contribute at least one ensemble member toward the 37 skillful simulations of the SIA decline (Table 1). In addition, our analysis suggests that odds for a skillful simulation increase for a larger model ensemble size. For example, among 11 CMIP6 models with more than 10 ensemble members, 10 (91%) CMIP6 models include at least 1 ensemble member reproducing the accelerated rate of Arctic sea ice loss. In contrast, only 7 out of 31 (22.5%) CMIP6 models with an ensemble size ranging from one to nine members exhibit similar skills. Given the complex and highly variable nature of Arctic sea ice trends, this highlights the need for a sufficiently large ensemble size in order to accurately capture the role of internal variability in these changes. Finally, considering the observed uncertainties, we find that 55 (19%) out of the 286 ensemble members from 27 (64%) CMIP6 models fall within that acceleration range from 1.44 to 2.09 (Fig. 6a and Table 1).

On the other hand, out of the total of 286 CMIP6 ensemble members examined here, SIA trends in 128 (45%) and 149 (52%) ensemble members fall outside the range of ±2σ from the observed mean P1 and P2 trend, respectively (Fig. 5). In addition, 48 (17%) ensemble members from the observed mean P1 and P2 trend signal in the P2 period. However, we note that for shorter periods (i.e., 18- versus 36-yr periods), one must recognize that the contribution of internal climate variability increases and makes the forced signal more difficult to detect (Kay et al. 2011; Swart et al. 2015).

Alternatively, model skill can also be evaluated based on ensemble-mean SIA anomalies using a paired t test. Based on such analysis, only 9 (21%) out of 42 CMIP6 models reasonably show the observed accelerated SIA trends, with the ratio of P2/P1 trends ranging between 1.99 and 2.83 (see Table 1). However, we find this approach less accurate given that 3 out of those 9 CMIP6 models (i.e., AWI-CM 1.1 LR, CAS-Esm2-0, and EC-Earth3) do not include a single skillful ensemble member that represents the observed SIA trend and acceleration given the criteria for statistical significance used above.

For the MMM trends, the P1 CMIP6 SIA trend is −0.22 × 10^6 km^2 decade^{-1}, with individual model ensemble means ranging from 0.19 × 10^6 km^2 decade^{-1} (E3SM 1.1 ECA) to −0.52 × 10^6 km^2 decade^{-1} (GFDL-CM4). The P2 trend is −0.56 × 10^6 km^2 decade^{-1}, with individual model ensemble means ranging from −0.02 × 10^6 km^2 decade^{-1} (NorESM2-LM) to more than twice the observed rate, −1.41 × 10^6 km^2 decade^{-1} (E3SM 1.0). Collectively, the CMIP6 MMM trends are statistically significant when compared to observations, but the accelerated trend is larger (i.e., a ratio of P2/P1 = 2.55) than the observed trend (Fig. 6a). All but six CMIP6 models (AWI-Esm1.1 LR, Cams-Csm 1.0, CNRM-CM6-1, GFDL-Esm4, NorEsm2-LM, NorEsm2-MM) show an accelerated rate of sea ice decline between periods P1 and P2. However, the rate of decline during P2 in 29 (69%) CMIP6 models is slower than that of observations, albeit notably faster (i.e., greater than −1.0 × 10^6 km^2 decade^{-1}) in CanEsm5, CanEsm5-CanOE, and E3SM 1.0.

It is important to note that the SIA anomaly linear trends and acceleration ratios in CMIP6 models are sensitive to how periods P1 and P2 are defined (Table S3). The absolute error from using an alternate later subperiod (e.g., breakpoint between 1999 and 2000; cf. Shu et al. 2020) for the observed SIA accelerated trend is only about 4.0%, while the models show greater sensitivity with an average error of about 37%; it ranges between 5.1% and 91.01%. Of 42 CMIP6 models examined here, six (CAM-Csm 1.0, EC-Earth3-Veg, GFDL-CM4, KIOST-ESM, MPI-ESM1.2-HAM, NorESM2-MM) show a stronger accelerated trend when we split the time series at the year 2000, which indicates these models simulate a stronger decline of sea ice later in the analysis period (Table S3). For 30 CMIP6 models, this occurs before our defined breakpoint between P1 and P2. As such, models with breakpoints in the early 1990s will include more negative trends during the defined P1 period than those with breakpoints later in the 1990s. The result is that the accelerated trend in these models tends to be lower (i.e., the upper left of the solid magenta line in Fig. 6a) than those with later breakpoints. Only six models (AWI-Esm1.1 LR, BCC-Csm 2 MR, EC-Earth3-Veg, KIOST-ESM, SAM0-UNICON, TaiESM1) have breakpoints occurring between 1996 and 2000, which best coincides with observations (Comiso et al. 2008; Stroeve et al. 2012b; Serreze and Stroeve 2015).

March SIA observations indicate no apparent acceleration of sea ice decline although the uncertainty is significant: declining sea ice from P1 (−0.45 km^2 decade^{-1}) to P2 (−0.39 × 10^6 km^2 decade^{-1}) (Fig. 6b). On the other hand, for September, observations show a substantially accelerated rate of SIA decline (P2/P1 = 3.1) from −0.36 × 10^6 km^2 decade^{-1} during P1 to −1.13 × 10^6 km^2 decade^{-1} during P2 (Fig. 6c). However, such a strong seasonal distinction is not present in the CMIP6 SIA MMM where the relative magnitude of the accelerated trend is the same (P2/P1 = 2.4) for both March and September, though the latter does have a larger magnitude of negative trends. It is clear that the September SIA decline drives the accelerated rate of loss observed in the overall ice cover (Stroeve and Notz 2018). However, this is not ubiquitous in CMIP6 models, of which many show a more dampened response meaning smaller September and larger March rates of SIA decline, possibly related to weak ice-albedo feedback.

A paired two-tailed t test confirms that the CMIP6 MMM SIV and PIOMAS trends are statistically different at p ≤ 0.10 between P1 and P2 periods (not shown). The CMIP6 SIV trend for P1 is −1.66 × 10^3 km^3 decade^{-1}, about 10% stronger than PIOMAS (−1.49 × 10^3 km^3), and for P2 the MMM trend is −3.55 × 10^3 km^3 decade^{-1}, which is about 25% weaker than PIOMAS (−4.74 × 10^3 km^3, Table 2). The results in the CMIP6 SIV MMM show an accelerated trend (P2/P1 = 2.1), which is two-thirds the rate of PIOMAS (P2/P1 = 3.2). On the other hand, the CMIP6 MMM accelerated trend is about twice that of RAS-M-G, which has no appreciable change in the SIV trend between P1 and P2 and can be interpreted as a lower bound (Table 2).

As was the case for SIA, large internal variability of the simulated SIV trends is exhibited between models and within a model’s ensemble members (Fig. 7). The model spread is larger when analyzing the shorter climate periods (i.e., P1 and P2)
FIG. 6. (a) SIA anomaly trends for the periods 1979–96 (P1 on the x axis) and 1997–2014 (P2 on the y axis) for the combined passive microwave observations and CMIP6 models with the total of 286 CMIP6 ensemble members (black dots) as well as the MMM. (b) March and (c) September trends of SIA anomaly. Error bars indicate a 95% confidence interval for each period. The solid magenta line illustrates the observed SIA acceleration ratio [slope of P2/P1 = (a) 1.78, (b) 0.86, and (c) 3.18], and the gray dashed line illustrates an acceleration ratio of 1.
rather than the full SIV historical simulation period (Fig. S2b). Trend uncertainty (i.e., error bars) for SIV is even larger than for SIA due to strong autocorrelation; SIV AR1 is greater than 0.85 for P1 and greater than 0.80 for P2 in all models (not shown). Both P1 and P2 CMIP6 SIV MMM and RASM-G trends are statistically different than zero, while the PIOMAS SIV P1 trend is not (Fig. 7). For PIOMAS, this is due in part to large trend uncertainty as a result of P1 AR1 = 0.95. Of the 286 CMIP6 ensemble members examined here, 247 (86%) members for P1 and 202 (71%) members for P2 have 2σ trend uncertainties within their respective CMIP6 SIV MMM 2σ trend uncertainties (Figs. 7a,b, respectively). Forty-two (15%) members from 19 CMIP6 models have a positive P1 trend, and, despite the larger negative trend signal of P2, seven members from five CMIP6 models show a positive P2 trend.

Fig. 7. SIV trends from (a) P1 (1979–96) and (b) P2 (1997–2014) for all individual model ensemble members, PIOMAS reanalysis, and RASM-G simulation. Error bars indicate adjusted 95% CI. The dark and light gray shadings indicate the adjusted 68% and 95% CI of CMIP6 SIV MMM, respectively [following Santer et al. (2008) and Stroeve et al. (2012a)].
While individual CMIP6 ensemble model mean SIV trends for P1 range between $2.16 \times 10^3$ km$^3$ decade$^{-1}$ (E3SM 1.1 ECA) and $-7.30 \times 10^3$ km$^3$ decade$^{-1}$ (E3SM 1.1), CMIP6 models exhibit a much faster decline ranging between $-0.30 \times 10^3$ km$^3$ decade$^{-1}$ (CIESM) and $-12.22 \times 10^3$ km$^3$ decade$^{-1}$ (E3SM 1.1 ECA) for P2 (Fig. 8a). All but seven CMIP6 models (AWI-ESM 1.1 LR, CESM-WACCM, CNRM-CM6-1, E3SM 1.0, GFDL-ESM4, MRI-ESM2, SAM0-UNICON) show an accelerated rate of sea ice decline between the periods P1 and P2, and therefore qualitatively match the PIOMAS SIV tendency (Table 2). The rate of decline during P2 in 30 (71%) CMIP6 models is slower than that of PIOMAS, albeit notably faster in nine models (i.e., greater than $-6.0 \times 10^3$ km$^3$ decade$^{-1}$): CanESM5-CanOE, E3SM 1.0, E3SM 1.1, E3SM 1.1 ECA, EC-Earth3-Veg, HadGEM3-GC31-LL, SAMO-UNICON, TaiESM1, and UKESM1.0-LL. The SIV accelerated trend is larger than PIOMAS for about 20% of CMIP6 models, while 29 (69%) CMIP6 models range between the accelerated trend values given by PIOMAS and RASM-G, 3.2 and 1.0, respectively.

The CMIP6 MMM SIV trends for both March and September are similar, showing P1 rates of sea ice decline around $-1.6 \times 10^3$ km$^3$ decade$^{-1}$ and P2 rates about $-3.5 \times 10^3$ km$^3$ decade$^{-1}$ (Figs. 8b,c). The same is true for the PIOMAS P1 trends for both March and September (approximately $-1.6 \times 10^3$ km$^2$ decade$^{-1}$), but the PIOMAS September P2 trend is 22% stronger than the March P2 trend ($-5.0 \times 10^3$ and $-4.1 \times 10^3$ km$^2$ decade$^{-1}$, respectively). SIV accelerated trends for the CMIP6 SIV MMM are about P2/P1 = 2.1 and show little seasonality, while PIOMAS accelerated trends range between P2/P1 = 2.8 and 3.5 (March and September, respectively). Thus, PIOMAS suggests some seasonal enhancement of the SIV decline during P2 while the CMIP6 MMM does not.

4. Discussion

Incorporating relevant observational metrics to constrain sea ice simulations is important to help identify models suited for further process-level analysis. Here we presented an analysis of the accelerated rate of sea ice decline between periods P1 and P2 (Figs. 6 and 8) to help better characterize CMIP6 models’ capability in representing the complex climate interactions between sea ice, ocean, and atmosphere. While SIA simulations are reasonably well constrained against passive microwave observations, simulations of SIV are not (Zygmuntowska et al. 2014). As such, our criteria for identifying a more accurate representation of historical SIV is less certain due to the lack of reliable observational data and thus requires discretion when interpreting results. Yet most CMIP6 models underrepresent the acceleration of sea ice decline in terms of both SIA and SIV.

a. The role of ocean heat transport in sea ice changes

Although recent Arctic sea ice changes are strongly linked to atmospheric warming via anthropogenic CO$_2$ (Notz and Stroeve 2016), a number of studies describe an increase in poleward OHT (Mayer et al. 2016, 2019; Tsubouchi et al. 2021) and its influence on sea ice decline (Årthun et al. 2012; Sandø et al. 2014; Li et al. 2017; Årthun et al. 2019; Asbjørnsen et al. 2020; Docquier and Koenigk 2021). Yet local ocean warming (via radiative fluxes) may also play an important role, especially in the western Arctic, since this heat remains available below the surface mixed layer through the winter and into the following summer (e.g., Steele et al. 2011). While a detailed analysis of possible causality of the upper ocean heat budget is beyond the scope of this paper, preliminary evidence from our study suggests increased northward OHT in CMIP6 models being closely linked to their simulations sharing the relatively realistic accelerated sea ice decline.

Only 17 of 42 CMIP6 models analyzed here provide the depth and zonally integrated variable meridional (poleward) OHT (i.e., hfbasin) from all ocean processes (e.g., resolved advective and diffusive heat transport; Griffies et al. 2016), which we use for a first-order examination of sea ice trends due to oceanic forcing (Fig. 9). We choose 80°N OHT to capture the influence of Atlantic Water into the Arctic through Fram Strait and the Barents Sea as well as to reduce the influence of local heating processes as the area above 80°N remains mostly ice covered in most models. A reasonably clear direct relationship is shown between positive northward OHT anomaly trends at 80°N and the magnitude of negative SIA anomaly trends in CanESM5, CanESM5-CanOE, EC-Earth3, EC-Earth3-Veg, HadGEM3-GC31-LL, HadGEM3-GC31-MM, IPSL-CM6A-LR, and UKESM1.0-LL (Fig. 9c). At the same time, the other models exhibit a relatively weaker SIA anomaly trend with a slightly negative to almost no OHT trend (Figs. 9a,c). For those models with strong OHT trends, except IPSL-CM6A-LR, the positive trend of poleward OHT is prevalent throughout the ocean between 65° and 80°N (Fig. S4). This appears to indicate that the magnitude of Atlantic OHT could play a role in explaining the ability of CMIP6 models to reproduce the observed decline in SIA.

Out of 42 CMIP6 models, only 10 models submitted both OHT output and sea ice mass change to the CMIP6 archive. Out of these 10, we selected 4 models (UKESM1.0-LL, HadGEM3-GC31-MM, MRI-ESM2.0, and NorESM2-MM) that represent a wide range of both SIA decline and OHT trends (see Table 1 and Fig. 9). Figure 10 demonstrates the critical contribution of top ice melt during summer, which drives an overall decline of sea ice in all four models. On the other hand, changes in net basal growth (fall and winter) and melt (summer) indicate a positive trend in sea ice mass change (i.e., a relative increase of sea ice thickness due to enhanced basal growth or reduced basal melting of overall thinning ice). The increasing trend of larger OHT corresponds to the stronger decline in frazil ice growth in models like UKESM1.0-LL and HadGEM3-GC31-MM (see Fig. 9 and Fig. S4). In contrast, NorESM2-MM has the lowest rate of sea ice decline yet shows a positive (increasing) trend in frazil ice growth because its OHT trend is negative in the Arctic (Fig. S4). This suggests an impact of OHT on sea ice production in sea ice mass change during the winter season (e.g., Sandø et al. 2014; Docquier and Koenigk 2021). However, the relationship ($r = -0.64, p \leq 0.05$) between OHT anomaly and SIV anomaly trends (Fig. 9d) is not as strong as shown for SIA (Fig. 9c; $r = -0.88, p < 0.05$). We hypothesize that loss of SIV in CMIP6 may include other...
Fig. 8. SIV anomaly trends for the periods 1979–96 (P1 on the x axis) and 1997–2014 (P2 on the y axis) for CMIP6 models, PIOMAS reanalysis, and RASM-G simulation. (b) March and (c) September trends of SIV anomaly. Error bars indicate a 95% confidence interval for each period. The solid magenta line illustrates the CMIP6 MMM SIV acceleration ratio [slope of P2/P1 = (a) 2.13, (b) 2.21, and (c) 1.96], the dotted magenta line illustrates the PIOMAS SIV acceleration ratio [slope of P2/P1 = (a) 3.18 (b) 2.71, and (c) 3.42], and the gray dashed line illustrates an acceleration ratio of 1. Note that E3SM 1.1 ECA is not shown because it is out of the axes range.
processes, such as ice growth-thickness feedback (Bitz and Roe 2004) and/or changes in ice motion and deformation (Schweiger et al. 2019). Interestingly, unlike SIA trends, models with stronger SIV anomaly trends tend to have thicker sea ice (Fig. S7).

b. Internal variability of sea ice decline in CMIP6 models

Internal variability of the climate must always be considered when comparing against observations, especially when evaluating trends on shorter time scales. Specifically, it is known that internal climate variability permits a range of possible outcomes of Arctic sea ice states, of which the observed state is but one realization (Swart et al. 2015; England et al. 2019). Such variability can account for as much as 50% of the September SIE trend in the pan-Arctic sea ice loss since 1979 (Kay et al. 2011; Stroeve et al. 2012b) and from 10% to 75% of regional SIC trends (England et al. 2019). For CMIP6 models, Shen et al. (2021) inferred that
FIG. 10. Spatially and temporally integrated (a),(d),(g),(j) top melt, (b),(e),(h),(k) bottom sea ice melt and basal growth, and (c),(f),(i),(l) frazil ice growth terms (kg; $\times 10^{15}$ kg) in (top to bottom) UKESM1-0-LL, HadGEM3-GC31-MM, MRI-ESM2-0, and NorESM2-MM, respectively. Slopes ($\times 10^{15}$ kg yr$^{-1}$) are calculated for the whole period (1979–2014).
approximately 22% ± 5% of the 1979–2014 September SIE trend can be attributed to model internal variability, assuming no bias in the model response to external forcing. While not quantified here, we demonstrated that analysis of shorter time series (e.g., less than 20 years in the analysis of accelerated rate of SIA/SIV decline) contributed to large trend uncertainties (Kay et al. 2011) in CMIP6. Although it is desirable to analyze a wide range of ensemble members, trend analysis is often based on ensemble means in which the influence of internal variability is removed. Therefore, it is not accurate to compare the model ensemble mean to the observed trend that reflects both forced response (climate warming) and natural variability. Moreover, it is difficult to measure what the true internal variability contribution is and how it compares to model internal variability (Wyburn-Powell et al. 2022). The next iteration of CMIP should have a longer overlapping period of historical experiments with a larger ensemble size and SIA observations to reduce trend uncertainty. Additionally, we intentionally analyze CMIP6 ensembles and the CMIP6 MMM to minimize the impact of internal model variability.

c. Uncertainties in evaluation of sea ice volume

Due to large uncertainties in estimated SIT observations and the corresponding SIV time series, we examine two SIV reanalyses, RASM-G and PIOMAS, alongside CMIP6 models. We note that the spatial and temporal SIT uncertainties vary widely within PIOMAS, with Schweiger et al. (2011) reporting for 1979–2010 a conservative trend uncertainty estimate of 1.0 × 103 km3 decade−1 based on three PIOMAS integration runs. The PIOMAS domainwide SIV bias relative to the SIT CDR (Lindsay 2010) has been estimated at −2.8 × 103 km3 for March and −1.5 × 103 km3 for October, or about 10% of the total SIV over the same period. In Fig. S5, we show the 12-month running means of PIOMAS and RASM-G SIV anomaly with linear trends. The RASM-G SIV mean is 21.0 × 103 km3 (Fig. 3b), and the linear trend is −2.52 × 103 km3 decade−1 for 1979–2014 (Table 2), which is 4% higher and 17% slower compared to the PIOMAS mean and trend, respectively. RASM-G simulated range of 15–27 × 103 km3 (Fig. 3a) is another expression of slightly thicker ice in P2, yet well-correlated SIV evolution compared to the PIOMAS range, with a 20% smaller standard deviation (Fig. 4b).

A short period of SIV observations from CryoSat-2 measurements of sea ice freeboard (October to April 2010–14) overlaps the CMIP6 historical period and offers an observational constraint for a portion of period P2, albeit too short for more than a qualitative interpretation here. Against the CryoSat-2 SIV time series, RASM-G shows a minimal bias and RMSE, PIOMAS shows a notable bias of −2.3 × 103 km3 and RMSE of 2.6 × 103 km3, and the CMIP6 SIV MMM splits the difference (i.e., bias of −1.0 × 103 km3 and RMSE of 1.9 × 103 km3; Figs. S6a,b,c). The PIOMAS negative SIV bias increases against higher observed SIV values (Fig. S6d; Laxon et al. 2013; Tilling et al. 2015); thus, the bias is most pronounced during the months of largest SIT (e.g., March–May). This likely contributes to a P2 trend that is too strong, resulting from overly deep troughs in the SIV anomaly during 2010–14 (Fig. 1b and Fig. S5).

The P1 SIV trend for RASM-G is twice as strong as that for PIOMAS, and it is larger than the CMIP6 MMM trend, while the P2 trend for RASM-G is 12% smaller than the MMM trend but 34% less than PIOMAS trend for P2 (Table 2). The overall result is only a slight increase in strength between P1 and P2 in RASM-G, which is half of the CMIP6 MMM acceleration ratio but close to 3 times smaller than the one for PIOMAS. Given the reasonable RASM-G skill in simulating both SIV time series and SIT spatial patterns (Watts et al. 2021), as well as the reported bias in PIOMAS SIV combined with uncertainty in its trend estimates, RASM-G offers an alternative to the PIOMAS perspective on the mean state and evolution of SIV in the Arctic. Further constraints on the historical evolution of SIT and SIV, i.e., a longer period of CryoSat-2 combined with Ice, Cloud and Land Elevation Satellite 2 (ICESat-2) observations, will assist future model intercomparison projects.

5. Conclusions

Most CMIP6 models simulate an accelerated rate of decline in SIA (86%) and SIV (88%) from the end of the twentieth century to the beginning of the twenty-first century. Twenty-eight CMIP6 models (67%) underestimate the SIA trend compared to observations (~0.54 × 106 km2 decade−1). Of 14 CMIP6 models that match or exceed the rate of SIA decline, 13 exhibit a positive SIA bias. Seventeen out of 42 CMIP6 models exhibit some skill since they have at least one ensemble member that reasonably represents the SIA decline and accelerated trends shown in the multialgorithm mean of the observed SIA. Among those 17 models, 10 out of 11 CMIP6 models have an ensemble size of more than 10 ensemble members, whereas the other 7 models come from 31 CMIP6 models with an ensemble size of less than 9 ensemble members. Alternatively, considering the observed uncertainties in each SIC algorithm, we find that 27 (64%) CMIP6 models fall within the range of observed acceleration between 1.44 and 2.09. Most CMIP6 models (83%) underestimate the accelerated rate of decline in SIV, as shown in PIOMAS reanalysis, although PIOMAS has been shown to underestimate thick ice (Fig. S6; Schweiger et al. 2011; Stroeve et al. 2014). On the other hand, the CMIP6 MMM accelerated rate in SIV (P2/P1 = 2.13) is 37% slower than the PIOMAS reanalysis. Overall, accelerated change (more than twice the P1 rates) of sea ice decline in CMIP6 MMM SIA and SIV are similar, but no accelerated decline is seen for the observed SIA in March. Given that SIV simulations are not well constrained by observations, identifying a more accurate representation of the historical SIV trend is a challenge. As shown in previous studies (Kay et al. 2011; Swart et al. 2015), internal variability is large among individual CMIP6 ensemble member simulations, particularly when examining shorter time periods (i.e., 18 versus 36 years).

We propose that further process-level examination of the CMIP6 model’s simulated OHT may advance understanding of its impact on sea ice decline and reduce uncertainties associated with ice–ocean feedback. An emergent constraint that we have diagnosed using 17 CMIP6 models
involves the interaction of the amount of ocean heat delivery and sea ice decline trends. Although some CMIP6 models show continuous sea ice reduction without prominent positive northward OHT trends, it is mainly driven by strong top ice melt during summer. The majority of models with a significant rate of sea ice decline are accompanied by a large increase in OHT into the Arctic. In particular, the sea ice mass change allows us to show that frazil ice growth increases (reduces) in those models with strong positive (negative) OHT trends (Fig. 10). Since our study is limited by the number of variables available in the CMIP6 archive, especially OHT and sea ice mass change, further investigation is needed to verify this emergent constraint relating trends in OHT to those in Arctic sea ice cover during a warming climate.

Acknowledgments. We would like to express our gratitude to the anonymous reviewers and the editor for their valuable feedback and insightful comments that helped to improve the quality and clarity of this manuscript significantly. 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 that support CMIP6 and ESGF. We acknowledge partial support from 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 the National Science Foundation (NSF) Arctic System Science (ARCSS) (to WM, 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-18-0496.1.


Schenker, N., and J. F. Gentleman, 2001: On judging the significance of differences by examining the overlap between...a229-2014.


