



## OPEN ACCESS

EDITED BY  
Jiankun Liu,  
Sun Yat-sen University, China

REVIEWED BY  
Zihan Wang,  
Polar Research Institute of China, China  
Sukun Cheng,  
University of Reading, United Kingdom

\*CORRESPONDENCE  
Marylou Athanase,  
✉ [marylou.athanase@awi.de](mailto:marylou.athanase@awi.de)

RECEIVED 11 November 2025  
REVISED 22 February 2026  
ACCEPTED 25 February 2026  
PUBLISHED 13 April 2026

## CITATION

Trivedi S, Sievers I, Athanase M,  
Sánchez-Benítez A and Semmler T  
(2026) Evaluating Arctic sea ice and  
snow thickness simulations:  
methodological insights from MOSAiC  
and CMIP6 data.  
*Front. Earth Sci.* 14:1744420.  
doi: 10.3389/feart.2026.1744420

## COPYRIGHT

© 2026 Trivedi, Sievers, Athanase,  
Sánchez-Benítez and Semmler. This is  
an open-access article distributed  
under the terms of the [Creative  
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).  
The use, distribution or reproduction in  
other forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which  
does not comply with these terms.

# Evaluating Arctic sea ice and snow thickness simulations: methodological insights from MOSAiC and CMIP6 data

Shreya Trivedi<sup>1,2</sup>, Imke Sievers<sup>3</sup>, Marylou Athanase<sup>4\*</sup>,  
Antonio Sánchez-Benítez<sup>4</sup> and Tido Semmler<sup>4,5</sup>

<sup>1</sup>Department of Geography, University of California, Los Angeles, United States, <sup>2</sup>NASA Goddard Space Flight Center, Greenbelt, MD, United States, <sup>3</sup>National Center for Klimaforskning, Danish Meteorological Institute, Copenhagen, Denmark, <sup>4</sup>Alfred Wegener Institute, Helmholtz-Centre for Polar and Marine Research, Bremerhaven, Germany, <sup>5</sup>Met Éireann, Dublin, Ireland

The Arctic sea ice cover and thickness have rapidly declined in recent years, with snow cover on sea ice playing a key role in driving this variability and trend. Arctic sea ice and snow properties also strongly regulate heat and momentum exchange between the ocean and atmosphere, making their accurate representation in climate models essential. Yet *in situ* observations of Arctic sea ice and snow thickness are scarce. The year-long Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) expedition provided valuable, high-resolution measurements of these properties, but this dataset is short-term and localized compared to the observational products typically used for model evaluation. We examine whether free-running climate model simulations can be meaningfully compared to point observations, such as those from MOSAiC, to assess model performance. To address this, we employ multiple methodological approaches to generate representative seasonal cycles of simulated snow and sea ice thickness: a standard 30-year climatology, two proxy-year methods (based on sea ice area (SIA) and the Arctic Oscillation (AO) index), atmospherically nudged simulations, and a Monte Carlo random-year benchmark. We find that the SIA-based proxy method performs comparably to the 30-year climatology. In contrast, the AO-based proxy method reduces bias relative to the SIA method for snow thickness comparisons. However, both methods nevertheless fail to accurately reproduce the amplitude of the observed snow thickness cycle, suggesting unresolved processes in models. Overall, these findings show potential for snow and sea ice evaluation against localized measurements and demonstrate that proxy methods can provide viable alternatives when nudging or direct temporal overlaps are unavailable. Finally, this study highlights the need for an improved representation of modeled sea ice and snow processes to enhance the next generation of global climate models.

## KEYWORDS

Arctic, climate models, cryosphere, MOSAiC, sea ice (Arctic), snow

## 1 Introduction

Sea ice and the overlying snow play a crucial role in the Arctic climate system (Parkinson and Cavalieri, 2008) through their insulating and reflective

properties (Maykut and Norbert, 1971). The rapid decline of Arctic sea ice in recent decades has largely affected the Arctic climate, its ecosystems, and the global atmosphere and ocean circulation (e.g., Stroeve et al., 2012; Hezel et al., 2012; Bhatt et al., 2014; Meier et al., 2014; Sévellec et al., 2017). Therefore, accurately simulating sea ice and snow properties and their variability in climate models is essential.

Several large-scale modes of internal climate variability influence Arctic sea ice variability and decline. Among these, the Arctic Oscillation (AO) is one of the most prominent (Webster et al., 2024) and is characterized by fluctuations in sea level pressure between the Arctic and mid-latitudes (Thompson and Wallace, 1998). AO-driven wind anomalies impact sea ice motion and export with seasonal lags (Rigor et al., 2002; Ogi and Wallace, 2010) and significantly shape sea ice and snow distributions (Wang and Ikeda, 2000; Webster et al., 2024). The AO phase determines the north–south displacement of the storm-steering jet stream, thus exerting a large influence on climate, weather, and sea ice variability across the Northern Hemisphere. Other influential modes include the quasi-biennial oscillation (QBO) in stratospheric winds, the El Niño–Southern Oscillation (ENSO) linked to equatorial Pacific temperature anomalies (Hu et al., 2016), and multidecadal modes such as the Atlantic Multidecadal Oscillation (AMO) and Pacific Decadal Oscillation (PDO) (Day et al., 2012). Additional contributors include sub-mesoscale eddies (Manucharyan and Thompson, 2017), mesoscale eddies (Gupta et al., 2020), intense winter storms in the Atlantic Arctic (Graham et al., 2019), and anomalous warm and humid air intrusions (Svensson et al., 2022). Ocean currents and heat transports also play roles in reducing sea ice area and volume, although their exact contributions remain uncertain (Docquier et al., 2021) and are beyond the scope of this study.

Global climate models (GCMs), particularly Coupled Model Intercomparison Project 6 (CMIP6) models (Eyring et al., 2016), have been widely assessed for their ability to simulate Arctic sea ice and snow, including their seasonality and overall volume distribution (Notz and SIMIP Community, 2020; Zhou et al., 2021; Watts et al., 2021; Xu and Li, 2023). Studies evaluating sea ice thickness (SIT) conclude that the simulated seasonality and trends agree well with reanalysis and satellite products, although large regional biases persist (Watts et al., 2021; Xu and Li, 2023). In contrast, Arctic snow thickness over sea ice has not been thoroughly evaluated, partly because Arctic-wide snow datasets have only recently become available (Zhou et al., 2021). Chen et al. (2021) compared a passive microwave snow thickness product with CMIP6 multi-model means and found that most models simulate a delayed snow maximum and snow thicknesses that are thinner than satellite-based estimates. To provide insights on sources of SIT and snow thickness biases in CMIP models, point-scale and process-focused studies are needed.

GCMs generally rely on simplified parameterizations for sea ice and snow on sea ice, typically using single-layer representations with constant snow density (e.g., Webster et al., 2021; Blanchard-Wrigglesworth et al., 2015; Hezel et al., 2012). This can introduce systematic biases, leading to considerable divergences in snow thickness trends across models (Webster et al., 2021). To address these biases, additional observations are essential. Airborne measurements from Operation IceBridge (MacGregor et al., 2021)

and reconstructions combining these data with reanalysis (Blanchard-Wrigglesworth et al., 2018) have advanced this field. Yet observational datasets still exhibit strong spatio-temporal heterogeneity in thickness, density, and water content (e.g., Webster et al., 2018), all of which affect the thermal properties of the sea ice and snowpack. This underscores the need for more comprehensive observational datasets to better evaluate and improve GCM performance.

The Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) is one of the most comprehensive *in situ* Arctic campaigns to date (Rabe et al., 2022; Rabe et al., 2024; Shupe et al., 2022; Nicolaus et al., 2022; Macfarlane et al., 2023a). In this study, we specifically use point-based observations that have not yet been assimilated into most climate models to test whether and how such localized measurements can be meaningfully compared to coarse-resolution model output. While satellite-derived thickness products exist (Petty et al., 2020; Laxon et al., 2013), they introduce their own uncertainties (e.g., freeboard retrieval errors, snow–ice interface assumptions, and limited temporal consistency), making MOSAiC uniquely suited for evaluating process-level model behavior despite its spatial constraints.

The year-long MOSAiC expedition provided simultaneous ocean, atmosphere, ice, and snow observations across a full annual cycle. From September 2019 to August 2020 (Rinke et al., 2021; Dethloff et al., 2022), the research vessel *Polarstern* (Alfred-Wegener-Institut, 2017) drifted with Arctic sea ice, enabling continuous measurements around the research vessel. The resulting snow and sea ice thickness data (Nicolaus et al., 2022; Wagner et al., 2022), collected over an area comparable to a typical GCM grid cell, provide an unparalleled reference for model evaluation. Indeed, satellite retrievals remain challenging without reliable ground observations or triple-collocation approaches (He et al., 2023), making *in situ* snow thickness data particularly valuable.

However, a direct comparison of point-based field data with coarse-resolution GCM output poses certain challenges.

1. Spatial differences arising from (i) discrepancies between localized MOSAiC measurements and grid-cell averages in GCMs and (ii) the spatial drift of the MOSAiC ice floe during the campaign.
2. Temporal differences stemming from the fact that MOSAiC represents a single year, capturing one realization of natural variability, while freely running climate models simulate generalized climatological conditions rather than capture specific years.

This study explores strategies to bridge these gaps and facilitate comparisons between *in situ* data and CMIP6 models. While such caveats are well known (Orlowsky and Seneviratne, 2014; Schutgens et al., 2017), our analysis shows that they do not preclude meaningful evaluation even with single-year, point-scale data, where first-order model deficiencies in SIT and snow are clearly detectable. This suggests that such comparisons are more robust and informative than often assumed, although future work is needed to disentangle whether this robustness applies equally to spatial sampling (point vs. grid cell) and temporal sampling (single year vs. climatology). Our study employs several methodological approaches to evaluate how well CMIP6 models reproduce the observed seasonal cycles of sea ice and snow thickness from MOSAiC and to compare the

performance of these approaches. Because free-running simulations do not reproduce the observed realization of natural variability, our complementary methods aim to minimize mismatches in sea ice variability, weather, and large-scale oscillations, isolating biases attributable to model physics or unresolved processes.

First, we propose and test two proxy-year selection methods: one based on sea ice area (SIA) and another based on an atmospheric criterion (the AO index). These methods intend to identify model years with large-scale characteristics similar to those during MOSAiC. Proxy-based selection is well established in paleoclimate research, particularly for temperature reconstructions (e.g., Miller et al., 2010; Burke et al., 2018). While a comprehensive understanding of all processes controlling Arctic sea ice and snow thickness in coupled climate models is beyond the scope of this study, we focus here on evaluating the representativeness of these proxies for model–observation comparison. We also employ nudged simulations, in which model atmospheric circulation is constrained to wind observations during the MOSAiC year (e.g., Sánchez-Benítez et al., 2022; Athanase et al., 2024a; Athanase et al., 2024b; Zhuo et al., 2025). These nudged simulations closely reproduce the observed conditions before and during MOSAiC (Pithan et al., 2022). Direct comparisons of the annual cycles of snow and ice thickness between nudged and free-running simulations thus help determine whether discrepancies primarily arise from mismatches in large-scale weather patterns or from limitations in model processes.

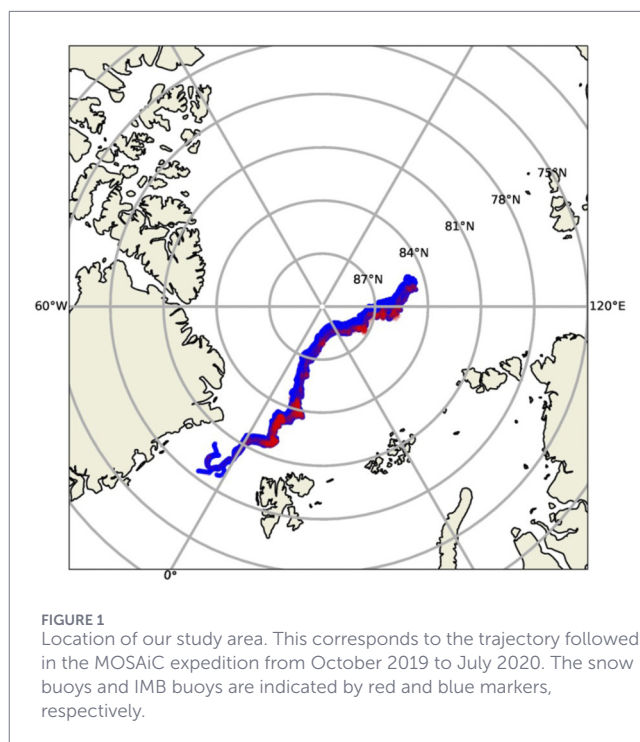
As not all GCMs have nudging capabilities, the proposed proxy methods can offer a broadly applicable alternative to nudged simulations, enabling comparatively direct model–observation comparisons of sea ice variables. To evaluate these methods, we compare them with (i) random-year selections generated with a Monte Carlo approach and (ii) the 30-year climatology of simulated sea ice variables centered on the MOSAiC year (2005–2034). Monte Carlo sampling, widely used to quantify statistical uncertainty (e.g., New and Hulme, 2000; Chen et al., 2022), provides a benchmark to assess whether our proxy and nudged methods yield results that differ significantly from random selection.

This study is organized as follows: Section 2 describes the observation and model datasets; Section 3 presents the proposed comparison methods; Section 4 reports the results; Section 5 discusses the findings and implications; and Section 6 provides concluding remarks and future outlook.

## 2 Datasets

### 2.1 Observations

Our study utilizes observational datasets from the Ocean and Sea Ice Satellite Application Facility (OSI-SAF) in the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) (Lavergne et al., 2019) for observed SIA over the pan-Arctic covering the time period from 1991 to 2022. We further use the observed AO index from the National Oceanic and Atmospheric Administration's (NOAA) Climate Prediction Center ([https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily\\_ao\\_index/ao.shtml](https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao.shtml)) (Sect. 3.1.2) over the MOSAiC time period.



*In situ* observations for snow thickness from the MOSAiC drift, used in the model–observation comparison, have been collected by sonar snow buoys (Nicolaus et al., 2021) and by sea ice mass balance (IMB) buoys (Lei et al., 2021). *In situ* observations for SIT are only from IMB buoys. In total, 32 buoys were considered in this study, which were deployed throughout the year in a 40-km radius around the MOSAiC ice floe (Nicolaus et al., 2022).

The snow buoys measure surface elevation changes relative to the day of deployment. The distance is measured by four ultrasonic sensors mounted on a square rack atop a 2.55 m pole. The snow thickness is derived from *in situ* snow thickness measurements at deployment and the measured elevation change. In addition to the elevation changes, the buoy also measures temperature and barometric pressure (Nicolaus et al., 2021). The preprocessed dataset from Nicolaus (2021) was used in this study. All obvious inconsistencies were removed during preprocessing, and the resulting data are provided, where available, in 1-h intervals. The snow buoys are indicated by red markers in Figure 1 and are partly covered by the IMB buoys displayed in blue.

The IMB buoys measure snow and SIT by exploiting their different thermal properties. To measure the SIT, an array of heating elements and temperature sensors is suspended from the top of the snowpack to the ocean. By cycling through cycles of heating and thermal diffusivity measurements, the surrounding medium can be determined to be ice, ocean, air, or snow (Jackson et al., 2013). This allows for a simultaneous measurement of both ice and snow thickness. The data used in this study were processed by Lei et al. (2021). The IMB buoys measure at a frequency of 1 day and have an accuracy of 0.02 m.

TABLE 1 Details of the specifications of the eight CMIP6 models used in the study. Nominal spatial resolution of each model component is in brackets.

Model name	Atmospheric model	Ocean model	Sea ice model	Reference
AWI-CM-1-1-MR or ECHAM6/FESOM	ECHAM6.3.04p1 [100 km]	FESOM1.4 [25 km]	FESIM 1.4 [25 km]	Semmler et al. (2018)
CESM2	CAM6 [100 km]	POP2 [100 km]	CICE5.1 [100 km]	Danabasoglu (2019)
CESM2-FV2	CAM6 [250 km]	POP2 [100 km]	CICE5.1 [100 km]	Danabasoglu (2019)
CESM2-WACCM	WACCM6 [100 km]	MAM4 [100 km]	CICE5.1 [100 km]	Danabasoglu (2019)
MPI-ESM1-2-HR	ECHAM6.3 [100 km]	MPIOM1.63 [50 km]	Unnamed thermodynamic (Semtner zero-layer) dynamic (Hibler 79) [50 km]	von Storch et al. (2017)
MPI-ESM1-2-LR	ECHAM6.3 [250 km]	MPIOM1.63 [250 km]	Unnamed thermodynamic (Semtner zero-layer) dynamic (Hibler 79) [250 km]	Wieners et al. (2019)
NorESM2-LM	CAM-OSLO [250 km]	MiCOM [100 km]	CICE [100 km]	Seland et al. (2019)
NorESM2-MM	CAM-OSLO [100 km]	MiCOM [100 km]	CICE [100 km]	Bentsen et al. (2019)

## 2.2 CMIP6 models

We analyze the monthly averaged variables over the period 1991–2050, expanding over the *historical* experiments (which ended in 2014) and the Shared Socioeconomic Pathways (SSP585) scenario of eight CMIP6 models (2015 and onward; Eyring et al., 2016; Notz et al., 2016) (Table 1). We focus on the “*sithick*” variable, representing simulated effective ice-floe thickness. We also incorporate “*sisdthick*,” representing the snow thickness. Additionally, for the proxy-year selection, we use the variables “*siarean*” and “*zg*” at 1000 hPa, representing the cumulative SIA over the Northern Hemisphere and the geopotential height, respectively. To avoid giving extra weight to models with multiple ensemble members, and to be consistent with previous CMIP6-based sea ice comparison studies (Notz and SIMIP Community, 2020; Roach et al., 2020), we considered the first ensemble member for each selected model (Table 1). This exploratory study tests our proposed methodology on eight CMIP6 models and a single ensemble member.

## 3 Methods

This exploratory study aims to advance the assessment of eight CMIP6 models in simulating the annual cycle of observed sea ice and snow thickness. To address the spatial and temporal challenges outlined in Section 1, we apply methodological solutions that mitigate, although do not fully eliminate, these limitations, without additional techniques such as downscaling. Instead, our approach focuses on reducing these mismatches within the capabilities of the available CMIP6 simulations.

- Model–observation spatial scale discrepancies are mitigated by averaging numerous snow and SIT buoy observations from MOSAiC. As GCMs provide a single mean value per grid cell, this averaging yields a comparable spatial scale, despite substantial sub-grid variability (Nicolaus et al., 2022). While not eliminating scale mismatches (as downscaling might), it mitigates them within this framework. To account for ice-floe drift, model values were collocated

with daily (nudged runs) or monthly (CMIP6 runs) observations (e.g., Athanase et al., 2019).

- To address the temporal discrepancies due to the CMIP6 models’ inability to simulate specific years, we propose three methods to enable meaningful year-to-year comparisons: (i) a simple proxy-year approach using the simulated sea ice area (SIA), (ii) another proxy-year selection using atmospheric criteria (AO index), and (iii) a nudging approach in which the atmospheric circulation of the GCM is relaxed toward the observed winds in the mid-troposphere.

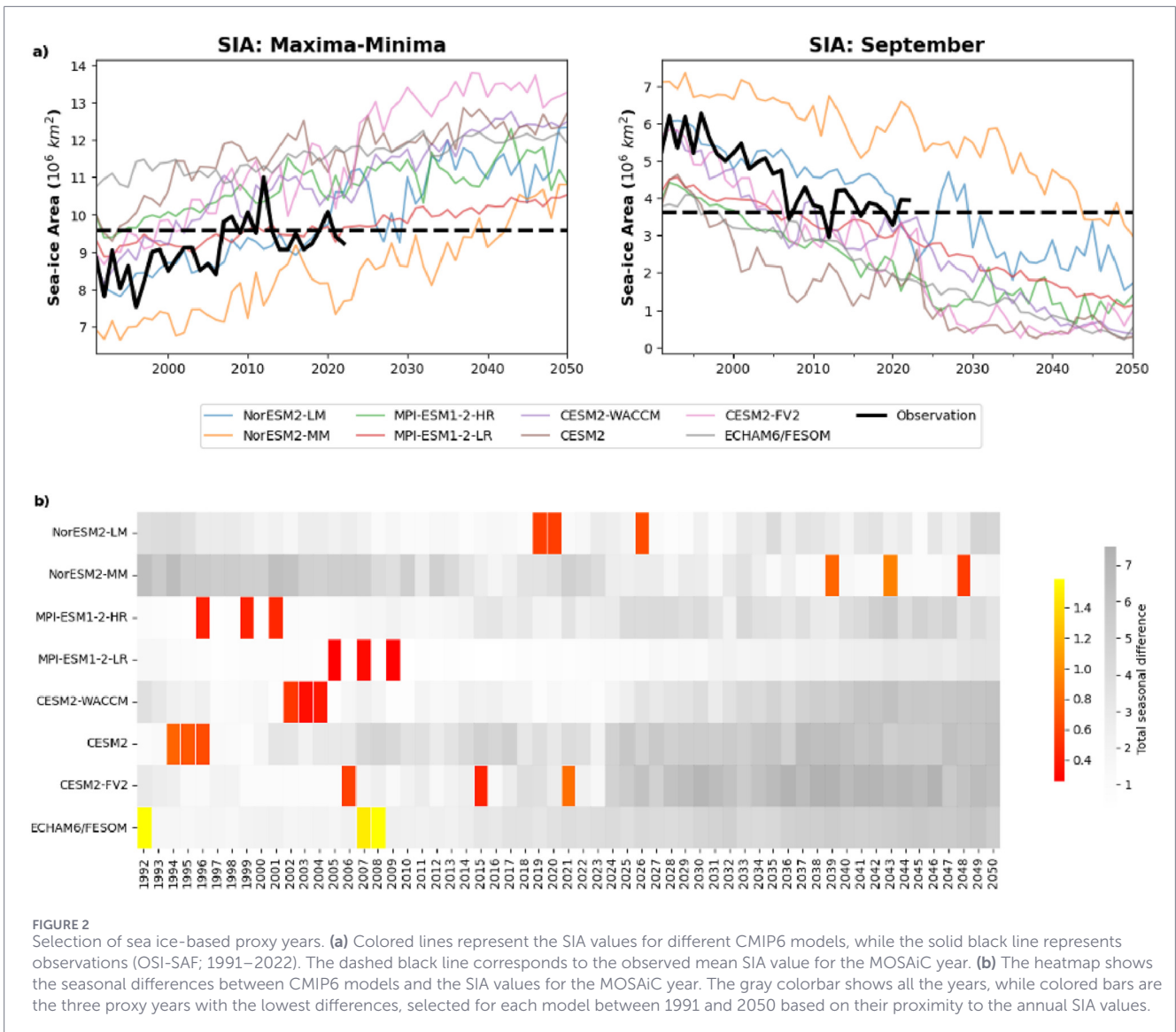
Our comparison methods assess whether simulated annual cycles of SIT and snow thickness improve under different selection strategies, either through targeted year selection based on sea ice or atmospheric indicators or via nudged simulations in which the atmosphere is constrained to observations. The MOSAiC campaign, designed to support climate model evaluation, included an extensive array of instruments deployed over an area comparable to a model grid cell (Shupe et al., 2022). Although these strategies cannot fully resolve scale mismatches or uncertainties, they offer a practical framework for evaluation in the absence of long-term *in situ* data.

### 3.1 Model–observation comparison methods

Freely running climate models generate different realizations of natural variability, making year-to-year comparisons with observations challenging. We propose proxy-year methods based on a single model realization, which are particularly useful when only one realization is available. These proxy years provide a generalizable framework to assess the influence of specific climate drivers (e.g., SIA anomalies or AO phases) across models. Their performance is compared with random-year benchmarks, as detailed below.

#### 3.1.1 Proxy years based on sea ice conditions

We propose selecting “proxy years” in CMIP6 models to align the simulated fields with conditions observed during the MOSAiC



**FIGURE 2** Selection of sea ice-based proxy years. **(a)** Colored lines represent the SIA values for different CMIP6 models, while the solid black line represents observations (OSI-SAF; 1991–2022). The dashed black line corresponds to the observed mean SIA value for the MOSAiC year. **(b)** The heatmap shows the seasonal differences between CMIP6 models and the SIA values for the MOSAiC year. The gray colorbar shows all the years, while colored bars are the three proxy years with the lowest differences, selected for each model between 1991 and 2050 based on their proximity to the annual SIA values.

campaign. Proxy years are first identified based on sea ice conditions analogous to those of the MOSAiC year. We use SIA rather than SIT due to the limited availability and high uncertainty of Arctic-wide SIT products, which rely on radar and laser altimetry and are affected by factors such as overlying snow, ice density, snow–ice interface, and retracking algorithms (Willatt et al., 2010; Kern et al., 2015; Xia and Xie, 2018; King et al., 2018; Landy et al., 2020; Sievers et al., 2024).

The SIA-based proxy years are selected based on two key metrics (Figure 2a).

1. Seasonal cycle amplitude,  $\Delta SIA = SIA \text{ (March)} - SIA \text{ (September)}$ , representing the first-year ice component.
2. September mean  $SIA_{Sep}$ , representing the contribution of multi-year ice.

For each model year  $i$ , we define an index of total seasonal difference  $D_i$  (Figure 2b) relative to the observational reference of

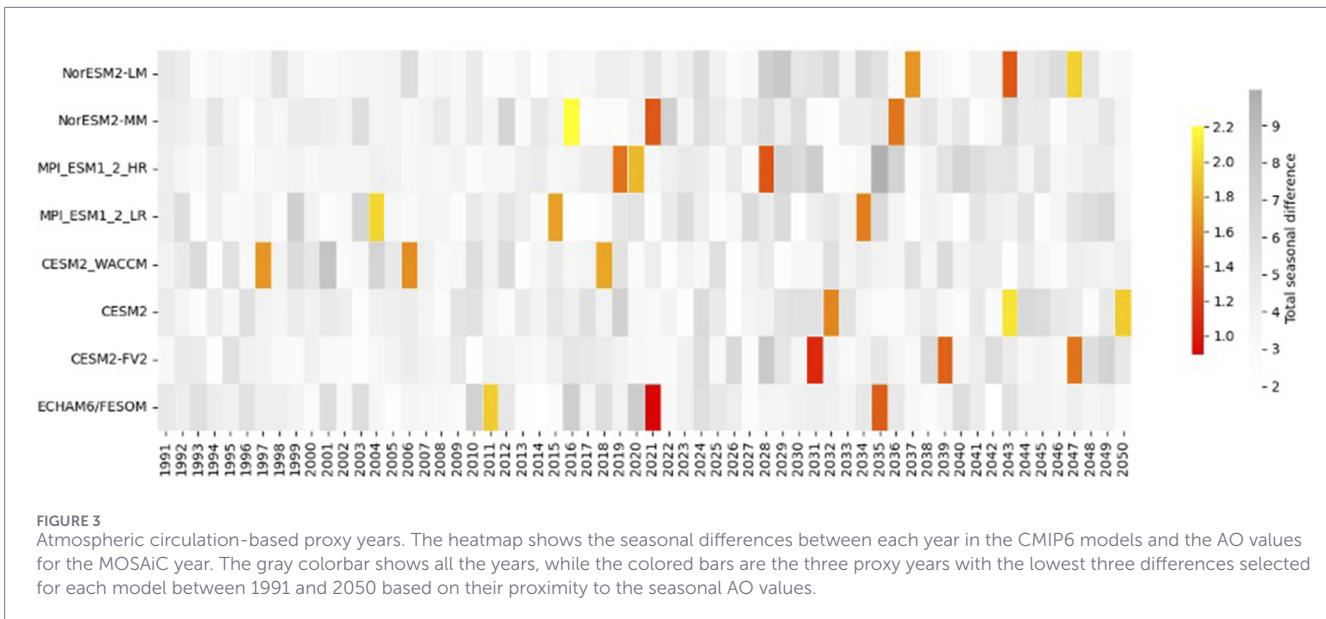
the MOSAiC year ( $ref$ ). It is computed as

$$D_i = |\Delta SIA_i - \Delta SIA_{ref}| + |SIA_{Sep,i} - SIA_{Sep,ref}|.$$

Here,  $\Delta SIA_i$  and  $SIA_{Sep,i}$  are the seasonal cycle amplitudes and the September mean SIA for a given model year  $i$ , respectively;  $\Delta SIA_{ref}$  and  $SIA_{Sep,ref}$  are the same metrics for the reference MOSAiC year. This index integrates differences in seasonal cycle amplitude  $\Delta SIA$  and in annual minimum  $SIA_{Sep}$  to indicate how well both MOSAiC conditions (both first-year and multi-year ice contributions) are represented in each model year. Proxy years are thus identified as those with the smallest absolute total seasonal difference  $D_i$  to most closely match the observed MOSAiC conditions.

### 3.1.2 Proxy years based on atmospheric (circulation) conditions

We also select proxy years based on the atmospheric circulation during the MOSAiC year using the AO index, which characterizes



mid-to-high latitude Northern Hemisphere circulation. For each model, the AO index is calculated as the normalized leading mode from empirical orthogonal function (EOF) analysis of geopotential height anomalies at 1000 hPa poleward of 20°N. During the MOSAiC year, the AO exhibited large shifts, ranging from a strongly negative index in November 2019 to a highly positive index in January–March 2020, marking it as an anomalous year in terms of AO behavior (Dethloff et al., 2022).

To account for this, we evaluate how closely seasonal AO index values in each model match observations. We therefore identify years with atmospheric conditions most similar to those during MOSAiC (calculated similarly to Figure 2B) and compare the observed and simulated AO index values for two key periods.

- Winter (January–March): corresponding to the strong positive AO phase during MOSAiC, and
- Preceding November: corresponding to the highly negative AO phase in late 2019.

For each model year  $i$ , the total AO difference relative to the MOSAiC reference year ( $ref$ ) is computed as

$$D_i^{AO} = |AO_{i,Nov} - AO_{ref,Nov}| + |AO_{i,JFM} - AO_{ref,JFM}|.$$

The 3 years with the smallest  $D_i^{AO}$  are selected as AO-based proxy years (Figure 3), representing the years whose simulated atmospheric circulation most closely matches the observed anomalous AO conditions of the MOSAiC period.

### 3.1.3 Nudged simulations

In addition to the proxy-year methods, we assess the potential of nudged simulations for meaningful model–observation comparisons of sea ice and snow cover. In these simulations, the observed wind evolution before and during MOSAiC is nudged (i.e., partially imposed) in coupled climate models. By reproducing the MOSAiC wind conditions, other model variables respond

freely but are closely guided by the nudged winds toward observed conditions (e.g., Sánchez-Benítez et al., 2022; Athanase et al., 2024a; Athanase et al., 2024b; Zhuo et al., 2025). Nudged simulations have proven effective for direct comparisons with Arctic atmospheric observations (Pithan et al., 2022). Here, we evaluate whether they can similarly support assessments of snow and sea ice thickness. This approach further helps determine whether discrepancies reflect mismatches in anomalous weather conditions or limitations in model representation of key processes.

Our nudged simulations employ two coupled climate models with spectral nudging: AWI-CM-1 (Sidorenko et al., 2015; Rackow et al., 2018) and AWI-CM-3 (Streffing et al., 2022) developed at the Alfred Wegener Institute. AWI-CM-1 couples ECHAM6.3.04p1 (Stevens, 2013) for the atmosphere with FESOM1.4 (Wang et al., 2014) for the ocean and sea ice, henceforth called ECHAM6/FESOM. The free-running ECHAM6/FESOM simulations contributed to CMIP6 (Semmler et al., 2020). AWI-CM-3 couples OpenIFS 43r3 (European Centre for Medium-Range Weather Forecasts ECMWF, 2017) for the atmosphere to FESOM2 (Danilov et al., 2017; Danilov et al., 2015), henceforth OpenIFS/FESOM2.

In both models, vorticity and divergence are spectrally nudged to ERA5 reanalysis data. Three ensemble members for each model are nudged from 1 January 2017, initialized from CMIP-type historical and subsequent SSP3-7.0 simulations, through August 2020, the end of the MOSAiC year. A 1-h relaxation timescale and spectral truncation of 20 (zonal wavenumbers for ECHAM6/FESOM; all wavenumbers for OpenIFS/FESOM2) yield the closest agreement with MOSAiC observations (Pithan et al., 2022). Only vertical levels between 700 hPa and 100 hPa are nudged, leaving the boundary layer, stratosphere, and other physical variables (temperature, surface pressure, humidity, clouds, precipitation, and sea ice) free to evolve. While ECHAM6/FESOM contributed to CMIP6, OpenIFS/FESOM2 is not included in proxy-year selection. Further details on the spectral nudging method are provided in Sánchez-Benítez et al. (2022).

## 3.2 Colocation to the MOSAiC drift trajectory

The selection of grid cells in the climate models is based on the MOSAiC trajectory. For each day or month, we determine the maximum north, south, west, and east extent of active snow and ice buoys and select the model grid cells covering this area. Due to varying resolutions and projections across CMIP6 models, the number of selected grid cells differs by month and model. Modeled grid-cell areas along the MOSAiC drift range from 999 km<sup>2</sup> (CESM2, July) to 3579 km<sup>2</sup> (MPI-ESM-1-2-LR, November). By comparison, the MOSAiC distributed network drifted within a ~30–40 km radius (~2,000–3,000 km<sup>2</sup>; Rabe et al., 2024), comparable to the model coverage. This approach thus accounts for variations in location and area (Section 3.3).

## 3.3 Processing of snow and ice observations

The snow buoys and Snow and Ice Mass Balance Array (SIMBA) buoys measure at different frequencies and resolutions. To get a thickness estimate comparable to the model grid cell values, each active buoy was averaged to a daily value. From this, the mean snow and SIT were calculated (shown in black in Figures 4a,b). From all buoys that are part of this averaging per day, the northernmost, southernmost, westernmost, and easternmost locations were determined to estimate the area over which the buoys were averaged. Figure 4c shows the number of active buoys per day. Overall, there were 13 snow buoys and 19 IMB buoys active throughout the period. The maximum number of instruments measuring snow thickness for 1 day was 23, while it was 17 for SIT.

The monthly mean was calculated from the daily averaged values in Figure 4. The locations for the model comparison were selected similarly as the locations for the daily mean, that is, by selecting the maximal east–west and north–south coverage of all observations considered within 1 month. The months of September 2019 and August 2020 were excluded from the mean and our further calculations because the observations did not cover the entire months, which would have led to biases in the monthly mean.

## 3.4 Benchmarks to evaluate comparison strategies

Our methodology employs two evaluation metrics to validate the comparison strategies. First, we use the Monte Carlo method to assess the reliability of the two proxy-year approaches (Section 3.1) and nudged simulations against random selections. We consider the eight CMIP6 models (Section 2.2) over 1991–2050. For each model, bootstrap resampling generates 10,000 iterations, randomly selecting 3 years per iteration to calculate the seasonal cycle. Results are then combined into a multi-model mean (MMM), producing numerous potential annual cycles of SIT and snow thickness. This allows evaluation of whether the proposed methods outperform random selection. Second, as an additional validation, we compute a 30-year climatology of SIT and snow thickness for each ensemble member, centered around the MOSAiC year (2005–2034). The

resulting MMM provides a baseline to assess the performance of the proxy-year methods and nudged simulations alongside Monte Carlo results.

## 4 Results

### 4.1 Comparisons for different *in situ* observations

To bridge the gap between model resolution and point observations, a large set of MOSAiC measurements is used to evaluate the simulated snow and sea ice thickness in GCMs. Snow thickness observations (Figure 4a) include both snow buoys and IMB buoys, averaged by weighting daily measurements according to the number of active buoys (Figure 4c). Some IMB snow measurements in summer 2020 were excluded. Snow buoys generally recorded thinner snow than IMB buoys during winter. From November to February, snow buoys measured a slight decrease in snow thickness, while IMB buoys measured a slight increase, likely reflecting a drop in active snow buoy sensors in January–February 2020 (red line in Figure 4c).

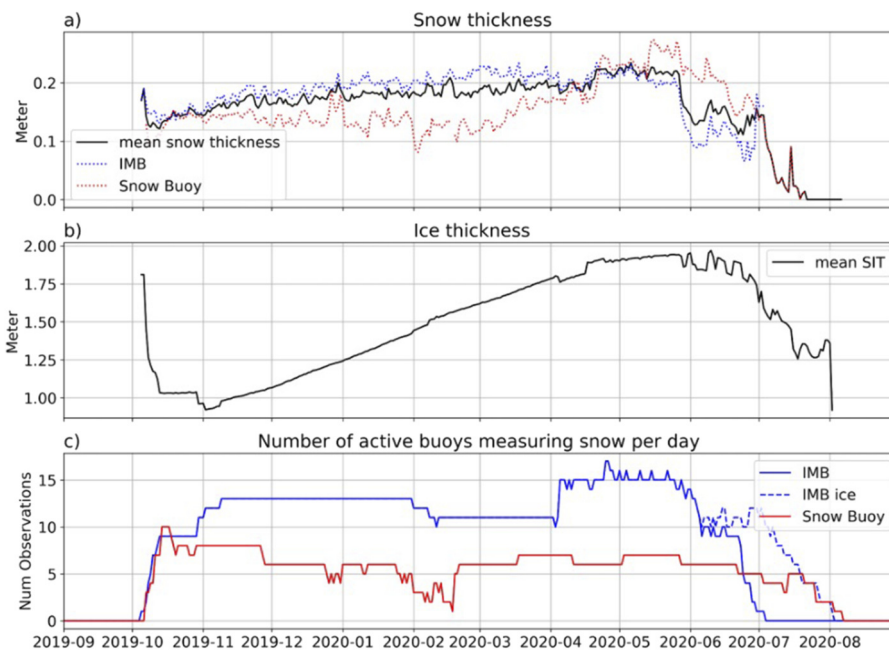
SIT observations (Figure 4b) are limited to IMB buoys, while snow buoys measure only snow. At the start and end of the observation period, mean SIT shows large variations due to the small number of active buoys. For example, one IMB measuring approximately 1.8 m SIT in early October strongly influences the mean; as additional buoys become active, this effect diminishes, producing an apparent decrease in the monthly mean SIT from October to November (Figures 5a, 6a, 8a). Both platforms require sufficient SIT to be deployed; thus, thin and fragile SIT conditions are underrepresented. This limitation should be considered when comparing observations with model output in the following sections, particularly in early autumn and summer.

### 4.2 Reproduction of annual cycles

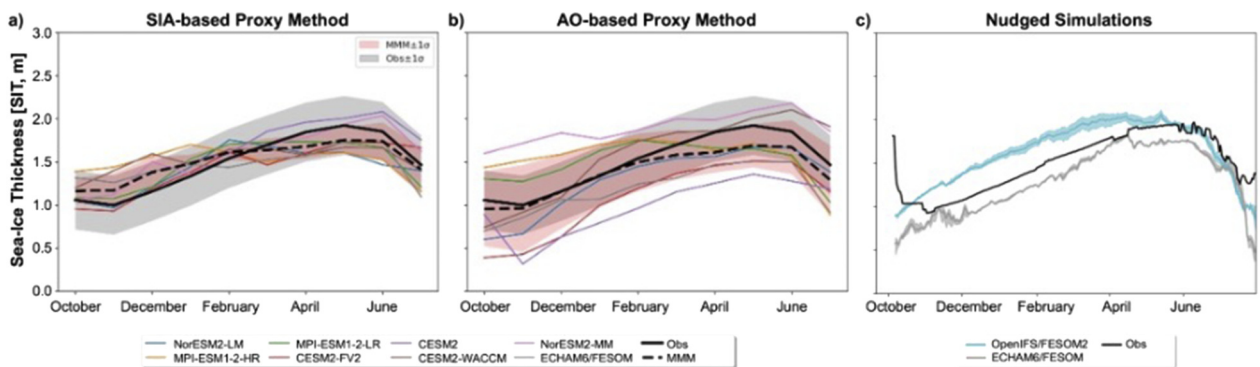
#### 4.2.1 Sea ice thickness

We examine the ability of CMIP6 models to reproduce the SIT annual cycles during MOSAiC, using proxy years based on SIA and AO criteria (Figure 5). Observations show a steady increase in SIT from November to May, followed by a decline, consistent with Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS) data (Chen et al., 2023). However, when employing SIA-based proxy years, CMIP6 models exhibit positive biases during cooler seasons (fall and winter) and underestimations during warmer seasons (spring and summer) (Figure 5a). Some models show an early SIT peak between January and March, reflecting rapid winter-spring buildup and excessive late-spring melt, as noted by Webster et al. (2021). Consequently, the MMM shows a nearly constant maximum SIT from February to June. The biases are higher for AO-based proxy years (Figure 5b), especially in cooler seasons. The MMM remains generally underestimated, with partial agreement with observations between November and January.

The inter-model spread, indicating uncertainty across model simulations, is calculated as the standard deviation of the annual SIT values averaged over the three selected years across all eight



**FIGURE 4**  
Time series of sea ice and snow parameters along the MOSAiC drift track. (a) Snow thickness (m), (b) sea ice thickness (m), and (c) number of active buoys. IMB buoys measure sea ice and snow thicknesses, while snow buoys measure only the snow thickness. Differences in the numbers of snow and ice observations from IMB in (c) are due to flawed snow observations.

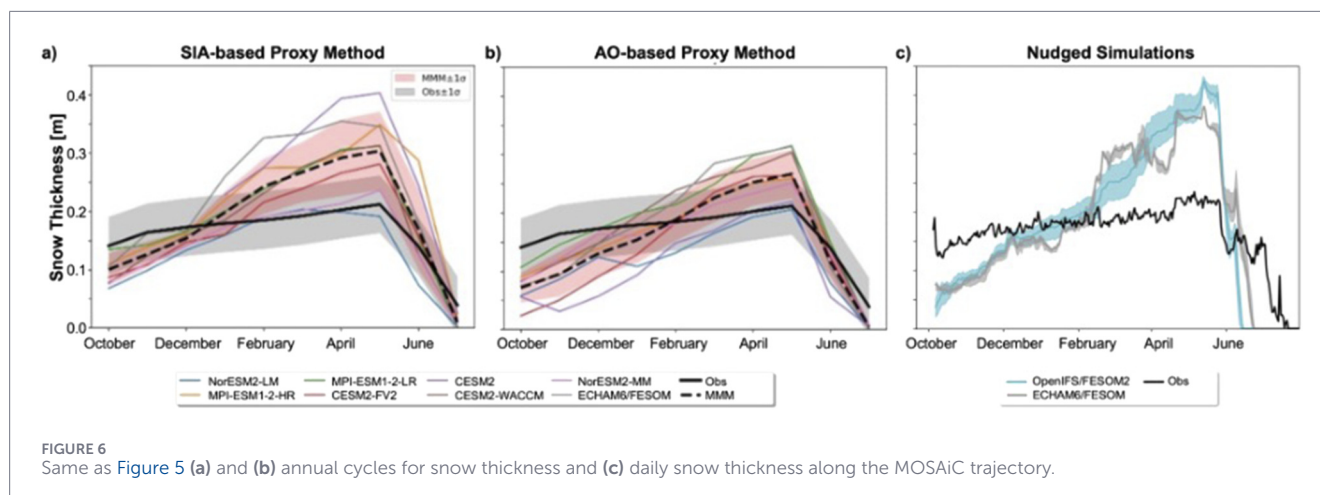


**FIGURE 5**  
Annual cycles for SIT averaged over the three selected years from (a) SIA-based and (b) AO-based methods. The parameters are averaged along the MOSAiC trajectory. The black solid line corresponds to the *in situ* observations, and the dashed black line shows the multi-model means (MMM). The colored lines represent different CMIP6 models. The shaded red areas represent  $\pm 1$  standard deviation for the MMM, and the gray shaded areas represent  $\pm 1$  standard deviation for the observations. Note that the observational standard deviation represents natural variability, while the MMM spread reflects inter-model variability; these quantities are shown for qualitative comparison only. (c) Daily SIT along the MOSAiC trajectory; the black line represents *in situ* observations, and the other lines show nudged simulations in which the evolution of winds observed during the MOSAiC year is imposed. Climate models used are ECHAM6/FESOM (blue) and OpenIFS/FESOM2 (grey). The shaded areas show the ensemble range for each set of coupled climate simulations; the thick lines show the ensemble mean.

CMIP6 models. This inter-model spread is slightly larger for AO-based selections, with 0.28 m against 0.22 m for SIA-based selections (Figure 5). Model spreads are relatively higher for most of the year but decrease briefly in January and February for the SIA-based method.

To complement the proxy-year analysis, we further utilize nudged simulations (see Section 3.1.3) to assess how explicitly constraining atmospheric circulation affects the simulated SIT

annual cycle and to identify remaining discrepancies arising from incomplete process representation in the models. The nudged simulations reproduce the MOSAiC SIT annual cycle reasonably well, with a mild underestimation in ECHAM6/FESOM and a slight overestimation for OpenIFS/FESOM2 (Figure 5c). Both capture a realistic SIT amplitude and seasonal signal, comparable to the MMM of the two proxy-year methods (from 0.6 m to 2 m; compare Figures 5a–c). The annual SIT maximum occurs in



late winter to spring and is slightly better represented in nudged simulations from April to June (Figures 5a,c). At the individual model level (e.g., when proxy-year selection is performed using only ECHAM6/FESOM), nudged simulations outperform proxy-based methods in simulating SIT (Supplementary Figure S1).

Overall, both proxy-year methods reasonably capture the timing and broad features of the SIT annual cycle, although some biases persist. The SIA-based method overestimates SIT in autumn and winter and underestimates it in spring and summer, whereas the AO-based selections show a general underestimation throughout the year. When compared with nudged simulations, the latter perform comparably to both proxy-year MMMs on monthly to seasonal timescales (Supplementary Figure S3) but additionally capture short-term variability, reproducing observed abrupt SIT changes linked to atmospheric events, such as the SIT drop in July. This underscores their added value in resolving transient events that proxy-based approaches cannot fully represent.

#### 4.2.2 Snow thickness

For snow thickness, model patterns and biases in the annual cycles are broadly consistent across both proxy methods. Snow thickness is underestimated from October to December (Figure 6), with AO-based proxies showing slightly higher negative biases in early winter (Figure 6b). From February to May, biases become progressively positive, as models simulate a gradual increase in snow thickness peaking in May, consistent with observations. The AO-based method shows slightly smaller positive biases during this period than the SIA-based method, likely reflecting the temporal window extending into the warmer SSP585 scenario (Figure 3). From June onward, both models and observations show a sharp decline, although modeled decreases are steeper, improving alignment with observations. Models also overestimate snow growth, likely contributing to the lower SIT growth rate, and are further discussed in Section 5.

The nudged simulations similarly show persistent biases: insufficient snow in fall and winter and excessive accumulation in spring (Figure 6c). Both nudged models overestimate snow during peak accumulation (March–May), particularly OpenIFS/FESOM2. Relative to the SIA-based proxy MMM, nudged simulations

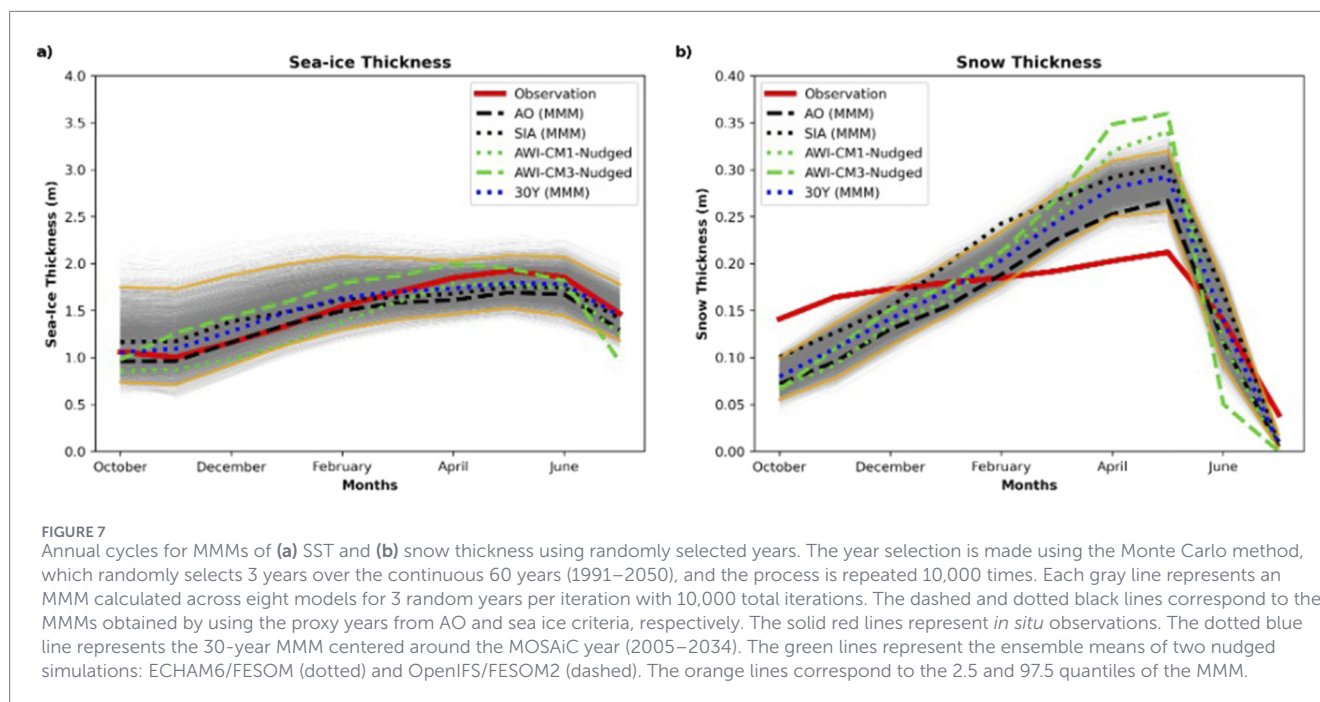
display similar performance, whereas the AO-based proxy MMM achieves better agreement, especially in spring and summer. At the individual model level (e.g., when AO-based proxy-year selection is applied solely to ECHAM6/FESOM), nudged simulations and AO-based proxies perform comparably in reproducing snow thickness (Supplementary Figure S1). As for SIT, nudged simulations match proxy-year MMMs on monthly timescales (Supplementary Figure S3) but additionally capture short-term variability, successfully reproducing observed snow peaks in late February, late April, and mid-June. This further highlights their skill in representing short-lived atmospheric events that affect snow depth.

Indeed, variations in precipitation and snowfall (as well as air temperature, see Pithan et al., 2022) in the nudged runs closely align with observations, with monthly accumulated values within 0.005 m of ERA5, except in May (0.015 m; Supplementary Figure S2). This indicates that snow thickness biases are not driven by atmospheric forcing, which is well captured in the nudged setup, but rather arise from inadequately represented surface processes such as snow advection, melt, and densification (see Section 5).

Annual cycles exhibit similar inter-model spreads of 0.09 m (SIA-based) and 0.08 m (AO-based), although the SIA-based method shows slightly higher disagreement between models from March to June. Spreads are smaller at the beginning and end of the cycle. Despite convergence in June, no method fully reproduces the observed magnitude of the snow thickness annual cycle for most models; only the Norwegian Earth System models (NorESM) show relatively good agreement. Overall, the proxy-year methods and nudged simulations perform less effectively for snow thickness than for SIT, emphasizing persistent deficiencies in the representation of key snow-related processes.

#### 4.3 Monte Carlo simulations and 30-year MMM for the annual cycles

Figure 7 shows the annual cycles of SIT and snow thickness derived from 10,000 Monte Carlo simulations, each selecting 3 years randomly. For SIT (Figure 7a), MMMs from both proxy-year criteria exhibit similar biases, aligning closely with the bootstrap distribution mean and the 30-year MMM (2005–2034). This



suggests that, for SIT, proxy-based methods offer limited advantage over climatology, though they better capture the seasonal evolution, closely matching observations (Figure 7a). Nudged simulations perform comparably, with ECHAM6/FESOM generally closer to observations and lower biases than OpenIFS/FESOM2. Notably, ECHAM6/FESOM approaches observed SIT from February to May before biases turn negative. While OpenIFS/FESOM2 overestimates SIT most of the year, aligning only partially between April and June, and at times falling outside the lower statistical limit, indicating statistically significant deviations from the random distribution.

For snow thickness (Figure 7b), models do not replicate any single year through bootstrapping that resembles observations, consistently underestimating in autumn and overestimating in May (Figure 7b). Both AO- and SIA-based proxy years display similar annual cycles but generally do not outperform the randomized Monte Carlo distribution. The AO-based method performs better, with lower biases than the SIA method and the 30-year MMM in February–May, confirming the value of atmospheric indices for variables strongly influenced by circulation. Nudged simulations consistently overestimate snow thickness, especially during March–May, with OpenIFS/FESOM2 showing the largest positive biases, exceeding observational and statistical limits, and underestimating in June–July. ECHAM6/FESOM performs slightly better but still exhibits substantial deviations.

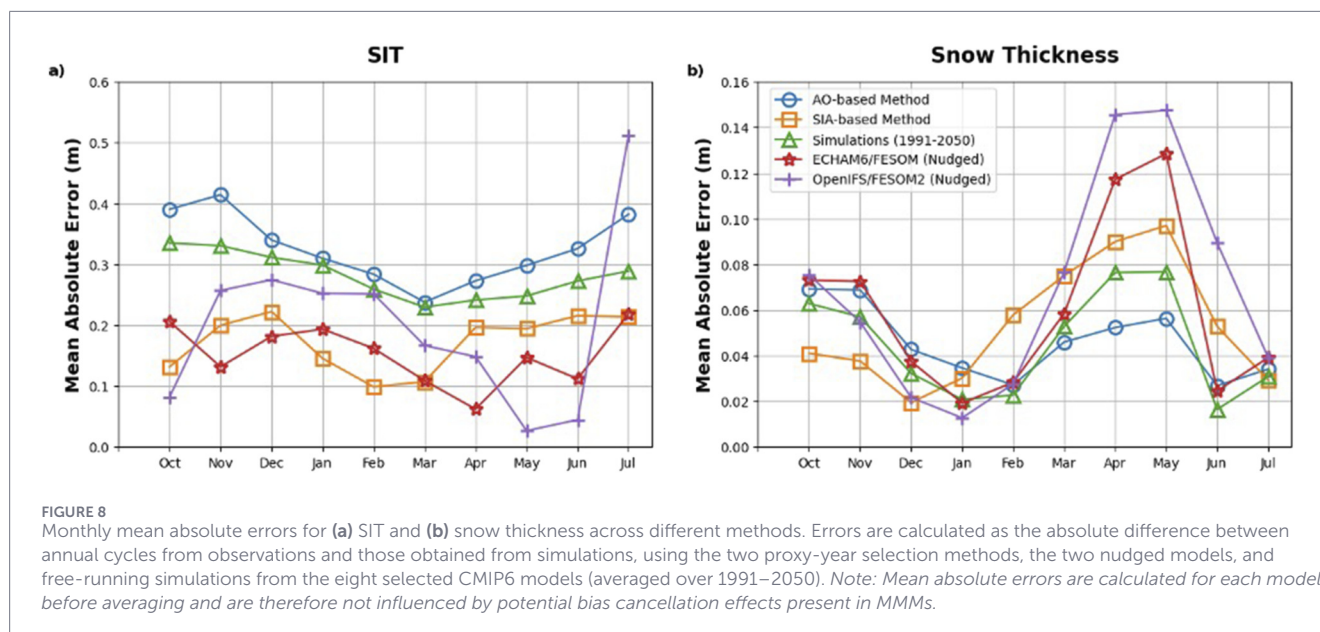
Overall, randomized distributions mostly include conditions completely different from the observations during the MOSAiC year. This suggests that, surprisingly, no matter which comparison strategy is applied, models consistently fail to reproduce MOSAiC snow thickness, indicating fundamental deficiencies in the representation of key processes, which remain underdeveloped or absent in current coupled climate models.

## 5 Discussion

This study demonstrates that CMIP6 models can be compared to and evaluated against localized, single-year datasets such as MOSAiC. To test how different methodological choices shape such evaluation, we employ two proxy-year selection methods: one based on sea ice area (SIA) and the other on the atmospheric Arctic Oscillation (AO) index (see methods), alongside nudged simulations from two GCMs, and conventional multi-year means. Surprisingly, all methods yield broadly consistent results, including a similar misrepresentation of the annual cycle, indicating that these proxy methods do not provide a clear advantage over multi-year means, even in the context of single-year observational comparisons.

Both proxy-year approaches partly mitigate, but cannot fully eliminate, spatial and temporal scale mismatches between observations and GCMs. Here, our aim is not to completely remove these differences but to test whether meaningful evaluation remains possible despite them. Nudged simulations, which constrain atmospheric circulation to MOSAiC conditions, allow us to separate discrepancies due to natural variability from those arising from underrepresented processes. Finally, our methods are evaluated using the 30-year MMM and Monte Carlo bootstrapping, in which the accuracy of annual cycles is tested over 10,000 randomly selected model years.

Across individual models, the MMM of SIA-based proxy years and nudged simulations produce lower errors in sea ice thickness (SIT) than both the AO-based proxy years MMM and the 1991–2050 MMM from free-running CMIP6 models. For snow thickness, AO-based proxy years have the smallest absolute errors (Figure 8b). Yet neither method accurately captures the annual cycle, and both exhibit large differences with observations, particularly in spring during the melting season. This confirms that snow-related processes are insufficiently represented in both CMIP6 and nudged models, resulting in low simulation accuracy across all methods



when compared to observations. In comparison, SIT is reasonably well represented in the models. As expected, the good performance of MMMs largely reflects temporal averaging and multi-model combinations, which smooth biases through compensating errors. When evaluating individual models, nudged simulations and AO-based proxy years perform comparatively better despite remaining large biases (Supplementary Figure S1).

Snow and sea ice performances across all methods reflect fundamental gaps in model physics. Despite accounting for spatial and temporal variations, biases along the MOSAiC drift persist for both SIT and snow thickness, particularly for snow thickness. SIT errors are generally reduced during winter months (Figure 8a), whereas snow thickness errors remain high (Figure 8b), with models overestimating the amplitude of annual cycles through exaggerated accumulation and melt. Nudged simulations reproduce monthly snowfall reasonably well (Supplementary Figure S2), indicating that overestimated accumulation arises from poorly represented snow advection, melt, densification (Warren et al., 1999; Macfarlane et al., 2023b), and reduction into leads (Clemens-Sewall et al., 2023), processes absent or underdeveloped in CMIP6 models. Furthermore, the underestimated snow layers in fall and winter may partly result from the observational setup, which required instruments to be deployed on level ice, thus accumulating thinner snow as wind redistributes snow toward ridges (Sturm et al., 2002). In contrast, the coarser model grids average over both level and ridged ice surfaces, contributing to the discrepancies. The underestimated snow thickness in March is followed by an underestimation in July, reflecting accelerated melt. This aligns with Arctic-wide analyses showing similar CMIP6 biases (Chen et al., 2021). However, some differences in the timing of the annual snow thickness maxima remain between the satellite-based data used in Chen et al. (2021) and the MOSAiC *in situ* measurements. This is likely because Chen et al. (2021) analyzed Arctic-wide averaged cycles of snow thickness, whereas we focus on point-scale measurements. Snow melt on sea ice is largely driven by the onset of shortwave radiation and albedo,

which exhibits high spatial variability (0.1–0.65) throughout the melt season (Perovich et al., 2002). Models cannot capture such localized processes, as they are constrained by coarse-resolution and albedo parameterizations tuned for sea ice extent (Hunke et al., 2010; Losch et al., 2010). This highlights the representation of snow on ice as a critical frontier for model improvement.

In contrast, SIT shows more robust agreement across methods, although seasonal variations in biases and inter-model spreads persist, with AO-based proxy years showing relatively higher spreads. Some of the SIT biases may arise as our study validates sea ice model output with IMB data. Such validation has inherent limitations as IMB buoys are only deployed on ice exceeding a minimum thickness and therefore do not reflect the variety of thicknesses found in one model grid cell (Olsen et al., 2026; Wever et al., 2021). Traditionally, comparisons rely on individual buoys, which are often installed on relatively thick and stable ice, being evaluated against grid-cell mean SIT, introducing representativeness errors. During the MOSAiC expedition, however, an array of buoys was deployed across spatial scales comparable to the typical model grid sizes and across ice with varying thickness. The deployment thicknesses in October/November ranged from approximately 30–175 cm, with 10 buoys placed on ice that was thinner than the average October thickness of approximately 100 cm, and six placed on ice that was thicker than the average. Several deployment reports further specify the installation of buoys on first-year ice (FYI) and multi-year ice (MYI). Despite this improved sampling, observational biases remain because an IMB cannot be installed on very thin ice and does not capture dynamically thickened features such as pressure ridges, which can build up to several meters thick (Valenti et al., 2020). Given that most buoys were deployed on thinner ice, the observations likely retain a thin bias relative to grid cell averages. Consequently, the model thin bias at the SIT maximum is likely underestimated.

Like the snow processes, SIT is also likely influenced by the unresolved snow processes mentioned herein. Many CMIP6

models assume uniform snow distribution across SIT categories, yet studies indicate that snow varies with SIT (Sturm et al., 2002; Liston, 2004; Castro-Morales et al., 2014). This misrepresentation can overly insulate thin ice, slowing ice growth (Figure 7), as seen in OpenIFS/FESOM2 during late winter and early spring (Figure 5c). Additional factors affecting SIT include sea ice albedo feedback (Kashiwase et al., 2017; Thackeray and Hall, 2019), increased heat loss during melting (Bitz and Roe, 2004; Hezel et al., 2012), complex ice–ocean feedback processes related to salinity and the sea-surface temperatures (Zhang et al., 2018; Goosse and Zunz, 2014), Atlantic Water biases (Khosravi et al., 2022), and regional atmospheric and surface wind inconsistencies in CMIP6 models (Crawford et al., 2023). The details of these biases are outside the scope of our study.

An important and somewhat unexpected result of our study is that, for SIT, the SIA-based proxy selection yields seasonal cycles and error characteristics that are nearly indistinguishable from those obtained using the 30-year climatology. While this may appear to reduce the added value of the proxy method for SIT, it is nevertheless informative. Conditioning on years with similar SIA does not substantially alter the simulated SIT seasonal cycle, and the resulting 30-year MMM lies within a relatively narrow range (Figure 7a). This suggests that the chosen SIA indicators, that is, the seasonal March–September difference and the September minimum, do not strongly constrain SIT evolution in the models analyzed here. Such behavior is consistent with a partial decoupling between SIA and thickness (Bocquet et al., 2024; Trivedi et al., 2025), whereby area variability does not necessarily translate into corresponding thickness adjustments. In this sense, the proxy approach functions as a diagnostic test of process coupling and representativeness: its similarity to the climatology indicates that conditioning on SIA alone does not meaningfully shift the modeled SIT cycle relative to the mean state.

Together, our results underscore two key points: first, SIT seasonal evolution is reasonably well simulated, and localized datasets like MOSAiC provide meaningful advantages over standard benchmarks; second, even basic comparison approaches can reveal persistent model shortcomings in snow thickness representation, highlighting the need for improved process representation in GCMs.

## 6 Conclusion

Evaluating the skill of global climate models (GCMs) using short-term, localized observational campaigns is challenging yet informative. Using MOSAiC observations, this study assesses three comparison strategies (two proxy-year approaches and an atmospherically nudged simulation) to benchmark model performance for two key Arctic parameters: sea ice and snow thickness.

Although no single comparison method fully reproduces MOSAiC observations, some strengths emerge across these approaches.

- For sea ice thickness (SIT) at the multi-model mean (MMM) level, the proxy method based on sea ice area (SIA) and the nudged runs capture seasonal amplitude most accurately.

- For snow thickness at the MMM level, none of the approaches accurately replicates the observed magnitude and seasonal cycle. However, the proxy method based on the Arctic Oscillation (AO) index achieves comparatively lower biases (despite persisting large discrepancies) and narrower inter-model spreads than the SIA-based proxy and the nudged simulations, especially during spring and summer. Nevertheless, even simple benchmarks such as a 30-year multi-model mean can reveal first-order model deficiencies.
- At the individual model level (e.g., ECHAM6/FESOM), the SIA- and AO-based proxy methods perform comparably to nudged simulations for SIT and snow thickness, respectively. This suggests that proxy-year selection can approximate some benefits of wind nudging when nudged simulations are unavailable.

Nonetheless, SIT is reasonably well captured across free-running and nudged model configurations, whereas snow thickness remains fundamentally misrepresented. No method reproduces the full observed annual cycle of snow thickness, highlighting persistent deficiencies in snow-related processes that may likely arise from overly uniform winter accumulation and excessively rapid snow and ice melt. Investigating the processes shaping snow and ice thickness, such as snow advection, compaction, redistribution, and albedo feedback, may offer insights into these biases. Expanding comparisons to other observational datasets, such as Operation IceBridge (Kurtz and Harbeck, 2015) and ICESat-2 (Kacimi and Kwok, 2022), would further test model performance under varying atmospheric and sea ice conditions.

Our study differs from previous sea ice comparison studies in the Arctic using CMIP6 models (Winkelbauer et al., 2024) in its selection of specific proxy years and the use of the unique MOSAiC trajectory for the comparison. This underscores the importance of localization in model–observation comparisons and suggests that CMIP models exhibit regional variations in ice formation processes that tend to average out on an Arctic-wide scale. Understanding these differences could help improve sea ice representation in GCMs. This study also suggests the value of relatively simple, experimental approaches for identifying dynamically consistent years in free-running CMIP6 outputs, even without direct temporal overlap with *in situ* observations. While not outperforming direct comparisons with simulations of the exact observational year, these methods are particularly valuable for evaluating historical simulations (ending in 2014) or future projections when overlap with observations is not possible. They also provide a practical alternative for research groups that lack the resources to perform computationally expensive nudged simulations. Therefore, this study contributes a flexible and accessible approach to model evaluation, helping bridge the gap between coupled climate models and field observations.

Overall, our study shows that even short, localized observational datasets under strongly anomalous conditions can provide insights for model evaluation, particularly for snow thickness. Improving the representation of key processes driving snow thickness, such as snowfall, thinning, compaction, redistribution mechanisms, and albedo, will be instrumental for the development of next-generation GCMs.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#); further inquiries can be directed to the corresponding author.

## Author contributions

ST: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. IS: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. MA: Data curation, Formal Analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing. AS-B: Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – review and editing. TS: Conceptualization, Formal Analysis, Investigation, Methodology, Supervision, Validation, Writing – original draft, Writing – review and editing.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. Imke Sievers was funded by the Danish State through the National Center for Climate Research (NCKF). Antonio Sanchez-Benitez was supported by funding from the Helmholtz Research Field Earth and Environment for the Innovation Pool Project SCENIC and ACTUATE. This work used resources of the Deutsche Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project ID ba1264.

## Acknowledgements

This work was developed as a part of the “Arctic Processes in CMIP6 Bootcamp” in October 2022 at Sømestationen, Denmark. The authors would like to thank the Danish Meteorological Institute,

## References

- Alfred-Wegener-Institut (2017). Helmholtz-Zentrum für Polar-und Meeresforschung: polar Research and Supply Vessel POLARSTERN Operated by the Alfred-Wegener-Institute. *J. Large-Scale Research Facilities* 3, A119. doi:10.17815/jlsrf-3-1631
- Athanase, M., Sennéchal, N., Garric, G., Koenig, Z., Boles, E., and Provost, C. (2019). New hydrographic measurements of the upper arctic western eurasian basin in 2017 reveal fresher mixed layer and shallower warm layer than 2005–2012 climatology. *J. Geophys. Res. Oceans* 124 (2), 1091–1114. doi:10.1029/2018jc014701
- Athanase, M., Sánchez-Benítez, A., Monfort, E., Jung, T., and Goessling, H. F. (2024a). How climate change intensified storm Boris' extreme rainfall, revealed by near-real-time storylines. *Commun. Earth Environ.* 5, 676. doi:10.1038/s43247-024-01847-0
- Athanase, M., Sánchez-Benítez, A., Goessling, H. F., Pithan, F., and Jung, T. (2024b). Projected amplification of summer marine heatwaves in a warming Northeast Pacific Ocean. *Commun. Earth Environ.* 5, 53. doi:10.1038/s43247-024-01212-1
- Bentsen, M., Oliviè, D. J. L., Seland, Ø., Toniazzo, T., Gjermundsen, A., Graff, L. S., et al. (2019). NCC NorESM2-MM model output prepared for CMIP6 CMIP historical. *Earth Syst. Grid Fed.* doi:10.22033/ESGF/CMIP6.8040
- Bhatt, U. S., Walker, D. A., Walsh, J. E., Carmack, E. C., Frey, K. E., Meier, W. N., et al. (2014). Implications of Arctic sea ice decline for the Earth system. *Annu. Rev. Environ. Resour.* 39 (1), 57–89. doi:10.1146/annurev-environ-122012-094357
- Bitz, C. M., and Roe, G. H. (2004). A mechanism for the high rate of sea ice thinning in the Arctic Ocean. *J. Clim.* 17 (18), 3623–3632. doi:10.1175/1520-0442(2004)017<3623:amfthr>2.0.co;2
- Blanchard-Wrigglesworth, E., Farrell, S. L., Newman, T., and Bitz, C. M. (2015). Snow cover on Arctic sea ice in observations and an Earth system Model. *Geophys. Res. Lett.* 42 (23), 10–342. doi:10.1002/2015gl066049
- Blanchard-Wrigglesworth, E., Webster, M. A., Farrell, S. L., and Bitz, C. M. (2018). Reconstruction of snow on arctic sea ice. *J. Geophys. Res. Oceans* 123 (5), 3588–3602. doi:10.1002/2017jc013364

NORP, and CLIVAR/CliC for organizing it. The authors also like to extend their gratitude to Tina Odaka for managing and equipping them with Pangeo EOSC Infrastructure (EGI-ACE).

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. We used large language models (LLMs) to assist with language refinement and copyediting of the manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feart.2026.1744420/full#supplementary-material>

- Bocquet, M., Fleury, S., Rémy, F., and Piras, F. (2024). Arctic and Antarctic sea ice thickness and volume changes from observations between 1994 and 2023. *J. Geophys. Res. Oceans* 129 (11), e2023JC020848. doi:10.1029/2023jc020848
- Burke, K. D., Williams, J. W., Chandler, M. A., Haywood, A. M., Lunt, D. J., and Otto-Bliesner, B. L. (2018). "Pliocene and Eocene provide best analogs for near-future climates," *Proc. Natl. Acad. Sci. U. S. A.* (Stanford, CA: National Academy of Sciences), 115 (52), 13288–13293. doi:10.1073/pnas.1809600115
- Castro-Morales, K., Kauker, F., Losch, M., Hendricks, S., Riemann-Campe, K., and Gerdes, R. (2014). Sensitivity of simulated Arctic sea ice to realistic ice thickness distributions and snow parameterizations. *J. Geophys. Res. Oceans* 119 (1), 559–571. doi:10.1002/2013jc009342
- Chen, S., Liu, J., Ding, Y., Zhang, Y., Cheng, X., and Hu, Y. (2021). Assessment of snow depth over Arctic sea ice in CMIP6 models using satellite data. *Adv. Atmos. Sci.* 38, 168–186. doi:10.1007/s00376-020-0213-5
- Chen, Y., Haywood, J., Wang, Y., Malavelle, F., Jordan, G., Partridge, D., et al. (2022). Machine learning reveals climate forcing from aerosols is dominated by increased cloud cover. *Nat. Geosci.* 15, 609–614. doi:10.1038/s41561-022-00991-6
- Chen, L., Wu, R., Shu, Q., Min, C., Yang, Q., and Han, B. (2023). The Arctic Sea ice thickness change in CMIP6's historical simulations. *Adv. Atmos. Sci.* 40, 1–13. doi:10.1007/s00376-022-1460-4
- Clemens-Sewall, D., Polashenski, C., Frey, M. M., Cox, C. J., Granskog, M. A., Macfarlane, A. R., et al. (2023). Snow loss into leads in Arctic sea ice: minimal in typical wintertime conditions, but high during a warm and windy snowfall event. *Geophys. Res. Lett.* 50 (12), e2023GL102816. doi:10.1029/2023gl102816
- Crawford, A. D., Rosenblum, E., Lukovich, J. V., and Stroeve, J. C. (2023). Sources of seasonal sea-ice bias for CMIP6 models in the Hudson Bay complex. *Ann. Glaciol.* 64, 1–18. doi:10.1017/aog.2023.42
- Danabasoglu, G. (2019). "NCAR CESM2 model output prepared for CMIP6 CMIP historical," Earth System Grid Federation. doi:10.22033/ESGF/CMIP6.7627
- Danilov, S., Wang, Q., Timmermann, R., Iakovlev, N., Sidorenko, D., Kimmritz, M., et al. (2015). Finite-element sea ice model (FESIM), version 2. *Geosci. Model Dev.* 8 (6), 1747–1761. doi:10.5194/gmd-8-1747-2015
- Danilov, S., Sidorenko, D., Wang, Q., and Jung, T. (2017). The finite-volume sea ice-ocean model (fesom2). *Geosci. Model Dev.* 10 (2), 765–789. doi:10.5194/gmd-10-765-2017
- Day, J. J., Hargreaves, J. C., Annan, J. D., and Abe-Ouchi, A. (2012). Sources of multi-decadal variability in arctic sea ice extent. *Environ. Res. Lett.* 7 (3), 034011. doi:10.1088/1748-9326/7/3/034011
- Dethloff, K., Maslowski, W., Hendricks, S., Lee, Y. J., Goessling, H. F., Krumpen, T., et al. (2022). Arctic sea ice anomalies during the MOSAiC winter 2019/20. *The Cryosphere* 16 (3), 981–1005. doi:10.5194/tc-16-981-2022
- Docquier, D., Koenigk, T., Fuentes-Franco, R., Karami, M. P., and Ruprich-Robert, Y. (2021). Impact of ocean heat transport on the Arctic sea-ice decline: a model study with EC-Earth3. *Clim. Dyn.* 56, 1407–1432. doi:10.1007/s00382-020-05540-8
- ECMWF (2017). IFS documentation CY43R3 – part III: dynamics and numerical procedures, no. 3 in IFS documentation. *ECMWF*. doi:10.21957/8l7mioid5m
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., et al. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* 9 (5), 1937–1958. doi:10.5194/gmd-9-1937-2016
- Goosse, H., and Zunz, V. (2014). Decadal trends in the Antarctic sea ice extent ultimately controlled by ice-ocean feedback. *Cryosphere* 8 (2), 453–470. doi:10.5194/tc-8-453-2014
- Graham, R. M., Itkin, P., Meyer, A., Sundfjord, A., Spreen, G., Smedsrud, L. H., et al. (2019). Winter storms accelerate the demise of sea ice in the Atlantic sector of the Arctic Ocean. *Sci. Rep.* 9 (1), 9222. doi:10.1038/s41598-019-45574-5
- Gupta, M., Marshall, J., Song, H., Campin, J. M., and Meneghello, G. (2020). Sea-ice melt driven by ice-ocean stresses on the mesoscale. *J. Geophys. Res. Oceans* 125 (11), e2020JC016404. doi:10.1029/2020jc016404
- He, L., Xue, B., Hui, F., and Cheng, X. (2023). Triple collocation-based merging of winter snow depth retrievals on arctic sea ice derived from three different algorithms using AMSR2. *IEEE Trans. Geoscience Remote Sens.* 62, 1–15. doi:10.1109/TGRS.2023.3290073
- Hezel, P. J., Zhang, X., Bitz, C. M., Kelly, B. P., and Massonnet, F. (2012). Projected decline in spring snow depth on arctic sea ice caused by progressively later autumn open ocean freeze-up this century. *Geophys. Res. Lett.* 39 (17), doi:10.1029/2012gl052794
- Hu, C., Yang, S., Wu, Q., Li, Z., Chen, J., Deng, K., et al. (2016). Shifting El Niño inhibits summer arctic warming and arctic sea-ice melting over the Canada basin. *Nat. Commun.* 7 (1), 11721. doi:10.1038/ncomms11721
- Hunke, E. C., Lipscomb, W. H., and Turner, A. K. (2010). Sea-ice models for climate study: retrospective and new directions. *J. Glaciol.* 56 (200), 1162–1172. doi:10.3189/002214311796406095
- Jackson, K., Wilkinson, J., Maksym, T., Meldrum, D., Beckers, J., Haas, C., et al. (2013). A novel and low-cost sea ice mass balance buoy. *J. Atmos. Ocean. Technol.* 30, 2676–2688. doi:10.1175/JTECH-D-13-00058.1
- Kacimi, S., and Kwok, R. (2022). Arctic snow depth, ice thickness, and volume from ICESat-2 and CryoSat-2: 2018–2021. *Geophys. Res. Lett.* 49 (5), e2021GL097448. doi:10.1029/2021gl097448
- Kashiwase, H., Ohshima, K. I., Nihashi, S., and Eicken, H. (2017). Evidence for ice-ocean albedo feedback in the Arctic Ocean shifting to a seasonal ice zone. *Sci. Rep.* 7 (1), 8170. doi:10.1038/s41598-017-08467-z
- Kern, S., Khvorostovsky, K., Skourup, H., Rinne, E., Parsakhoo, Z. S., Djepa, V., et al. (2015). The impact of snow depth, snow density and ice density on sea ice thickness retrieval from satellite radar altimetry: results from the ESA-CCI sea ice ECV project round robin exercise. *Cryosphere* 9, 37–52. doi:10.5194/tc-9-37-2015
- Khosravi, N., Wang, Q., Koldunov, N., Hinrichs, C., Semmler, T., Danilov, S., et al. (2022). The Arctic Ocean in CMIP6 models: biases and projected changes in temperature and salinity. *Earth's Future* 10 (2), e2021EF002282. doi:10.1029/2021ef002282
- King, J., Skourup, H., Hvidegaard, S. M., Rösel, A., Gerland, S., Spreen, G., et al. (2018). Comparison of freeboard retrieval and ice thickness calculation from ALS, ASIRAS, and CryoSat-2 in the Norwegian arctic to field measurements made during the N-ICE2015 expedition. *J. Geophys. Res.-Oceans* 123, 1123–1141. doi:10.1002/2017jc013233
- Kurtz, N., and Harbeck, J. (2015). Operation IceBridge sea ice freeboard, snow depth, and thickness data products manual, version 2 processing. Available online at: <https://nsidc.org/sites/default/files/idcsi4-icebridge-products-manual-version2-june-2015.pdf> (Accessed on April 2, 2026).
- Landy, J. C., Petty, A. A., Tsamados, M., and Stroeve, J. C. (2020). Sea ice roughness overlooked as a key source of uncertainty in CryoSat-2 ice freeboard retrievals. *J. Geophys. Res.-Oceans* 125, e2019JC015820. doi:10.1029/2019JC015820
- Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., et al. (2019). Version 2 of the EUMETSAT OSI SAF and ESA CCI sea ice concentration climate data records. *Cryosphere* 13 (1), 49–78. doi:10.5194/tc-13-49-2019
- Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., et al. (2013). CryoSat-2 estimates of arctic sea ice thickness and volume. *Geophys. Res. Lett.* 40 (4), 732–737. doi:10.1002/grl.50193
- Lei, R., Cheng, B., Hoppmann, M., and Zuo, G. (2021). *Snow depth and sea ice thickness derived from the measurements of SIMBA buoys deployed in the Arctic Ocean during the legs 1a, 1, and 3 of the MOSAiC campaign in 2019-2020*. PANGAEA. 938244. doi:10.1594/PANGAEA
- Liston, G. E. (2004). Representing subgrid snow cover heterogeneities in regional and global models. *J. Clim.* 17, 1381–1397. doi:10.1175/1520-0442(2004)017<1381:RSSCHI>2.0.CO;2
- Losch, M., Menemenlis, D., Campin, J. M., Heimbach, P., and Hill, C. (2010). On the formulation of sea-ice models. Part 1: effects of different solver implementations and parameterizations. *Ocean. Model.* 33 (1-2), 129–144. doi:10.1016/j.ocemod.2009.12.008
- Macfarlane, A. R., Schneebeli, M., Dacic, R., Tavri, A., Immerz, A., Polashenski, C., et al. (2023a). A database of snow on sea ice in the central arctic collected during the MOSAiC expedition. *Sci. Data* 10 (1), 398. doi:10.1038/s41597-023-02273-1
- Macfarlane, A. R., Löwe, H., Gimenes, L., Wagner, D. N., Dacic, R., Ottersberg, R., et al. (2023b). Temporospatial variability of snow's thermal conductivity on arctic sea ice. *Cryosphere* 17 (12), 5417–5434. doi:10.5194/tc-17-5417-2023
- MacGregor, J. A., Boisvert, L. N., Medley, B., Petty, A. A., Harbeck, J. P., Bell, R. E., et al. (2021). *The scientific legacy of NASA's operation IceBridge*.
- Manucharyan, G. E., and Thompson, A. F. (2017). Submesoscale sea ice-ocean interactions in marginal ice zones. *J. Geophys. Res. Oceans* 122 (12), 9455–9475. doi:10.1002/2017jc012895
- Maykut, G. A., and Norbert, U. (1971). Some results from a time-dependent thermodynamic model of sea ice. *J. Geophys. Res.* 76 (6), 1550–1575. doi:10.1029/jc076i006p01550
- Meier, W. N., Hovelsrud, G. K., van Oort, B. E., Key, J. R., Kovacs, K. M., Michel, C., et al. (2014). Arctic sea ice in transformation: a review of recent observed changes and impacts on biology and human activity. *Rev. Geophys.* 52 (3), 185–217. doi:10.1002/2013rg000431
- Miller, G. H., Alley, R. B., Brigham-Grette, J., Fitzpatrick, J. J., Polyak, L., Serreze, M. C., et al. (2010). Arctic amplification: can the past constrain the future? *Quat. Sci. Rev.* 29 (15-16), 1779–1790. doi:10.1016/j.quascirev.2010.02.008
- New, M., and Hulme, M. (2000). Representing uncertainty in climate change scenarios: a Monte-Carlo approach. *Integr. Assess.* 1, 203–213. doi:10.1023/A:1019144202120
- Nicolaus, M., (2021). *Snow height on sea ice, meteorological conditions and drift of sea ice from autonomous snow buoys during MOSAiC 2019/20*. PANGAEA.
- Nicolaus, M., Hoppmann, M., Arndt, S., Hendricks, S., Katlein, C., Nicolaus, A., et al. (2021). Snow depth and air temperature seasonality on sea ice derived from snow buoy measurements. *Front. Mar. Sci.* 8, 655446. doi:10.3389/fmars.2021.655446

- Nicolaus, M., Perovich, D. K., Spreen, G., Granskog, M. A., von Albedyll, L., Angelopoulos, M., et al. (2022). Overview of the MOSAiC expedition: snow and sea ice. *Elem. Sci. Anth* 10 (1), 000046. doi:10.1525/elementa.2021.000046
- Notz, D., and SIMIP, Community (2020). Arctic sea ice in CMIP6. *Geophys. Res. Lett.* 47 (10), e2019GL086749. doi:10.1029/2019gl086749
- Notz, D., Jahn, A., Holland, M., Hunke, E., Massonnet, F., Stroeve, J., et al. (2016). The CMIP6 Sea-Ice Model Intercomparison Project (SIMIP): understanding sea ice through climate-model simulations. *Geosci. Model Dev.* 9 (9), 3427–3446. doi:10.5194/gmd-9-3427-2016
- Ogi, M., and Wallace, J. M. (2010). The role of summer surface wind anomalies in the summer Arctic sea ice extent in 2010 and 2011. *Geophys. Res. Lett.* 39 (9). doi:10.1029/2012GL051330
- Olsen, I. B. L., Skourup, H., Sallila, H., Hendricks, S., Hansen, R. M. F., Kern, S., et al. (2026). A first approach towards dual-hemisphere sea ice reference measurements from multiple data sources repurposed for evaluation and product intercomparison of satellite altimetry. *Earth Syst. Sci. Data* 18 (3), 2469–2505. doi:10.5194/essd-18-2469-2026
- Orlowsky, B., and Seneviratne, S. I. (2014). On the spatial representativeness of temporal dynamics at European weather stations. *Int. J. Climatol.* 34 (10), 3154–3160. doi:10.1002/joc.3903
- Parkinson, C. L., and Cavalieri, D. J. (2008). Arctic sea ice variability and trends, 1979–2006. *J. Geophys. Res. Oceans* 113, C7. doi:10.1029/2007jc004558
- Perovich, D. K., Grenfell, T. C., Light, B., and Hobbs, P. V. (2002). Seasonal evolution of the albedo of multiyear Arctic sea ice. *J. Geophys. Res. Oceans* 107 (C10), SHE–20. doi:10.1029/2000jc000438
- Petty, A. A., Kurtz, N. T., Kwok, R., Markus, T., and Neumann, T. A. (2020). Winter Arctic sea ice thickness from ICESat-2 freeboards. *J. Geophys. Res. Oceans* 125 (5), e2019JC015764. doi:10.1029/2019jc015764
- Pithan, F., Athanase, M., Dahlke, S., Sánchez-Benítez, A., Shupe, M. D., Sledd, A., et al. (2022). *Nudging allows direct evaluation of coupled climate models with in-situ observations: a case study from the MOSAiC expedition.* Göttingen: Copernicus Publications, 1–23.
- Rabe, B., Heuzé, C., Regnery, J., Aksenov, Y., Allerholt, J., Athanase, M., et al. (2022). Overview of the MOSAiC expedition: physical oceanography. *Elem. Sci. Anth* 10 (1), 00062. doi:10.1525/elementa.2021.00062
- Rabe, B., Cox, C. J., Fang, Y. C., Goessling, H., Granskog, M. A., Hoppmann, M., et al. (2024). The MOSAiC distributed Network: observing the coupled Arctic system with multidisciplinary, coordinated platforms. *Elem. Sci. Anthropocene* 12 (1), 00103. doi:10.1525/elementa.2023.00103
- Rackow, T., Goessling, H. F., Jung, T., Sidorenko, D., Semmler, T., Barbi, D., et al. (2018). Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part II: climate variability. *Clim. Dynamics* 50, 2369–2394. doi:10.1007/s00382-016-3192-6
- Rigor, I. G., Wallace, J. M., and Colony, R. L. (2002). Response of sea ice to the arctic oscillation. *J. Clim.* 15 (18), 2648–2663. doi:10.1175/1520-0442(2002)015<2648:rositt>2.0.co;2
- Rinke, A., Cassano, J. J., Cassano, E. N., Jaiser, R., and Handorf, D. (2021). Meteorological conditions during the MOSAiC expedition: normal or anomalous? *Elem. Sci. Anth* 9 (1), 00023. doi:10.1525/elementa.2021.00023
- Roach, L. A., Dörr, J., Holmes, C. R., Massonnet, F., Blockley, E. W., Notz, D., et al. (2020). Antarctic sea ice area in CMIP6. *Geophys. Res. Lett.* 47 (9), e2019GL086729. doi:10.1029/2019gl086729
- Sánchez-Benítez, A., Goessling, H., Pithan, F., Semmler, T., and Jung, T. (2022). The July 2019 European heat wave in a warmer climate: storyline scenarios with a coupled model using spectral nudging. *J. Clim.* 35 (8), 2373–2390. doi:10.1175/JCLI-D-21-0573.1
- Schutgens, N., Tsyro, S., Gryspeerdt, E., Goto, D., Weigum, N., Schulz, M., et al. (2017). On the spatio-temporal representativeness of observations. *Atmos. Chem. Phys.* 17 (16), 9761–9780. doi:10.5194/acp-17-9761-2017
- Seland, Ø., Bentsen, M., Olivieri, D. J. L., Toniazzo, T., Ada, G., Seland, L., et al. (2019). NCC NorESM2-LM model output prepared for CMIP6 CMIP historical. *Earth Syst. Grid Fed.* doi:10.22033/ESGF/CMIP6.8036
- Semmler, T., Danilov, S., Rackow, T., Sidorenko, D., Barbi, D., Hegewald, J., et al. (2018). AWI AWI-CM1.1MR model output prepared for CMIP6 CMIP. *Earth Syst. Grid Fed.* doi:10.22033/ESGF/CMIP6.359
- Semmler, T., Danilov, S., Gierz, P., Goessling, H. F., Hegewald, J., Hinrichs, C., et al. (2020). Simulations for CMIP6 with the AWI climate model AWI-CM1-1. *J. Adv. Model. Earth Syst.* 12 (9), e2019MS002009. doi:10.1029/2019ms002009
- Sévellec, F., Fedorov, A. V., and Liu, W. (2017). Arctic sea-ice decline weakens the Atlantic meridional overturning circulation. *Nat. Clim. Change* 7 (8), 604–610. doi:10.1038/nclimate3353
- Shupe, M. D., Rex, M., Blomquist, B., Persson, P. O. G., Schmale, J., Uttal, T., et al. (2022). Overview of the MOSAiC expedition: atmosphere. *Elem. Sci. Anth* 10 (1), 00060. doi:10.1525/elementa.2021.00060
- Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., et al. (2015). Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part I: model formulation and mean climate. *Clim. Dyn.* 44, 757–780. doi:10.1007/s00382-014-2290-6
- Sievers, I., Skourup, H., and Rasmussen, T. A. S. (2024). Impact assessment of snow thickness, sea ice density and water density in CryoSat-2-derived sea ice thickness. *Cryosphere* 18 (12), 5985–6004. doi:10.5194/tc-18-5985-2024
- Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., et al. (2013). Coauthors: atmospheric component of the MPI-M Earth system model: ECHAM6. *J. Adv. Model. Earth Syst.* 5, 146–172. doi:10.1002/jame.20015
- Streffing, J., Sidorenko, D., Semmler, T., Zampieri, L., Scholz, P., Andrés-Martínez, M., et al. (2022). AWI-CM3 coupled climate model: description and evaluation experiments for a prototype post-CMIP6 model. *Geosci. Model Dev.* 15 (16), 6399–6427. doi:10.5194/gmd-15-6399-2022
- Stroeve, J. C., Serreze, M. C., Holland, M. M., Kay, J. E., Malanik, J., and Barrett, A. P. (2012). The Arctic's rapidly shrinking sea ice cover: a research synthesis. *Clim. Change* 110, 1005–1027. doi:10.1007/s10584-011-0101-1
- Sturm, M., Holmgren, J., and Perovich, D. K. (2002). Winter snow cover on the sea ice of the Arctic Ocean at the surface heat budget of the Arctic Ocean (SHEBA): temporal evolution and spatial variability. *J. Geophys. Res. Oceans* 107 (C10), SHE–23. doi:10.1029/2000jc000400
- Svensson, G., Murto, S., Shupe, M. D., Pithan, F., Magnusson, L., Day, J. J., et al. (2022). Warm air intrusions reaching the MOSAiC expedition in April 2020—The YOPP targeted observing period (TOP). *Elem. Sci. Anthropocene* 11 (1), 00016. doi:10.1525/elementa.2023.00016
- Thackeray, C. W., and Hall, A. (2019). An emergent constraint on future arctic sea-ice albedo feedback. *Nat. Clim. Change* 9 (12), 972–978. doi:10.1038/s41558-019-0619-1
- Thompson, D. W., and Wallace, J. M. (1998). The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.* 25 (9), 1297–1300. doi:10.1029/98GL00950
- Trivedi, S., Hobbs, W., and Raphael, M. (2025). An assessment of antarctic sea-ice thickness in CMIP6 simulations with comparison to the satellite-based observations and reanalyses. *Cryosphere* 19, 6771–6790. doi:10.5194/tc-19-6771-2025
- Valenti, V., Mahoney, A., and Metzger, A. (2020). A probabilistic description of pressure ridge width, spacing, and keel depth for the Chukchi and beaufort seas based on IPS and ADCP observations. *Cold Regions Sci. Technol.* 182, 103171. doi:10.1016/j.coldregions.2020.103171
- von Storch, J.-S., Putrasahan, D., Lohmann, K., Gutjahr, O., Jungclaus, J., Bittner, M., et al. (2017). MPI-M MPIESM1.2-HR model output prepared for CMIP6 HighResMIP. *Earth Syst. Grid Fed.* doi:10.22033/ESGF/CMIP6.762
- Wagner, D. N., Shupe, M. D., Cox, C., Persson, O. G., Uttal, T., Frey, M. M., et al. (2022). Snowfall and snow accumulation during the MOSAiC winter and spring seasons. *Cryosphere* 16 (6), 2373–2402. doi:10.5194/tc-16-2373-2022
- Wang, Q., Danilov, S., Sidorenko, D., Timmermann, R., Wekerle, C., Wang, X., et al. (2014). The finite element sea ice-ocean model (FESOM) v. 1.4: formulation of an ocean general circulation model. *Geosci. Model Dev.* 7 (2), 663–693. doi:10.5194/gmd-7-663-2014
- Wang, J., and Ikeda, M. (2000). Arctic oscillation and Arctic sea-ice oscillation. *Geophys. Res. Lett.* 27 (9), 1287–1290. doi:10.1029/1999GL002389
- Warren, S. G., Rigor, I. G., Untersteiner, N., Radionov, V. F., Bryazgin, N. N., Aleksandrov, Y. I., et al. (1999). Snow depth on arctic sea ice. *J. Clim.* 12, 1814–1829. doi:10.1175/1520-0442(1999)012<1814:SDOASI>2.0.CO;2
- Watts, M., Maslowski, W., Lee, Y. J., Kinney, J. C., and Osinski, R. (2021). A spatial evaluation of arctic sea ice and regional limitations in CMIP6 historical simulations. *J. Clim.* 34 (15), 6399–6420. doi:10.1175/jcli-d-20-0491.1
- Webster, M., Gerland, S., Holland, M., Hunke, E., Kwok, R., Lecomte, O., et al. (2018). Snow in the changing sea-ice systems. *Nat. Clim. Change* 8 (11), 946–953. doi:10.1038/s41558-018-0286-7
- Webster, M. A., DuVivier, A. K., Holland, M. M., and Bailey, D. A. (2021). Snow on arctic sea ice in a warming climate as simulated in CESM. *J. Geophys. Res. Oceans* 126 (1), e2020JC016308. doi:10.1029/2020JC016308
- Webster, M. A., Riihelä, A., Kacimi, S., Ballinger, T. J., Blanchard-Wrigglesworth, E., Parker, C. L., et al. (2024). Summer snow on arctic sea ice modulated by the arctic oscillation. *Nat. Geosci.* 17, 995–1002. doi:10.1038/s41561-024-01525-y
- Wever, N., Leonard, K., Maksym, T., White, S., Proksch, M., and Lenaerts, J. T. M. (2021). Spatially distributed simulations of the effect of snow on mass balance and flooding of antarctic sea ice. *J. Glaciol.* 67 (266), 1055–1073. doi:10.1017/jog.2021.54

- Wieners, K. H., Giorgetta, M., Jungclaus, J., Reick, C., Esch, M., and Bittner, M., (2019). MPI-M MPI-ESM1.2-LR model output prepared for CMIP6 CMIP historical *Earth System Grid Federation*. doi:10.22033/ESGF/CMIP6.6595
- Willatt, R. C., Giles, K. A., Laxon, S. W., Stone-Drake, L., and Worby, A. P. (2010). Field investigations of Ku-Band radar penetration into snow cover on antarctic sea ice. *IEEE Trans. Geoscience Remote Sens.* 48 (1), 365–372. doi:10.1109/TGRS.2009.2028237
- Winkelbauer, S., Mayer, M., and Haimberger, L. (2024). Validation of key arctic energy and water budget components in CMIP6. *Clim. Dyn.* 62, 3891–3926. doi:10.1007/s00382-024-07105-5
- Xia, W., and Xie, H. (2018). Assessing three waveform retrackerers on sea ice freeboard retrieval from Cryosat-2 using operation IceBridge airborne altimetry datasets. *Remote Sens. Environ.* 204, 456–471. doi:10.1016/j.rse.2017.10.010
- Xu, M., and Li, J. (2023). Assessment of sea ice thickness simulations in the CMIP6 models with CICE components. *Front. Mar. Sci.* 10, 1223772. doi:10.3389/fmars.2023.1223772
- Zhang, J., Stegall, S. T., and Zhang, X. (2018). Wind–sea surface temperature–sea ice relationship in the chukchi–beaufort seas during autumn. *Environ. Res. Lett.* 13 (3), 034008. doi:10.1088/1748-9326/aa9adb
- Zhou, Lu, Stroeve, J., Xu, S., Petty, A., Tilling, R., Winstrup, M., et al. (2021). Inter-comparison of snow depth over arctic sea ice from reanalysis reconstructions and satellite retrieval. *Cryosphere* 15 (1), 345–367. doi:10.5194/tc-15-345-2021
- Zhuo, W., Sánchez-Benitez, A., Athanase, M., Jung, T., Yao, Y., and Goessling, H. (2025). Storylines reveal contrasting thermodynamic effects of climate change on 2020/21 East Asian cold extremes. *Npj Climate Atmospheric Science* 8, 169. doi:10.1038/s41612-025-01031-x