JGR Atmospheres

Key Points:

- A cloudy state, without strong low-level stability (LLS), is often observed over winter Arctic sea ice but is absent in most Coupled Model Intercomparison Project (CMIP6) models
- CMIP6 models show a realistic representation of the dependence of LLS on near-surface air temperature and wind speed
- Observations show a decreasing trend in Arctic winter LLS, which CMIP6 models project will continue under warming

Correspondence to:

A. Duffey,
alistair.duffey.21@ucl.ac.uk

Author Contributions:

Conceptualization: Robbie Mallett, Jonathan Day, Felix Pithan
Data curation: Robbie Mallett
Formal analysis: Alistair Duffey, Robbie Mallett, Victoria R. Dutch, Julia Steckling, Antoine Hermant
Investigation: Alistair Duffey, Robbie Mallett, Victoria R. Dutch, Julia Steckling, Antoine Hermant
Methodology: Alistair Duffey, Robbie Mallett, Julia Steckling, Antoine Hermant, Jonathan Day, Felix Pithan
Project administration: Alistair Duffey, Jonathan Day, Felix Pithan

Representation of Arctic Winter Atmospheric Boundary Layer Stability Over Sea Ice in CMIP6 Models

Alistair Duffey[1] , Robbie Mallett[2], Victoria R. Dutch[3,4], Julia Steckling[5,6] , Antoine Hermant[7,8] , Jonathan Day[9] , and Felix Pithan[10]

[1]Centre for Polar Observation and Modelling, Earth Sciences, University College London, London, UK, [2]Earth Observation Group, Department of Physics and Technology, UiT The Arctic University of Norway, Tromso, Norway, [3]Department of Geography and Environmental Sciences, Northumbria University, Newcastle, UK, [4]School of Environmental Sciences, University of East Anglia, Norfolk, UK, [5]School of Integrated Climate and Earth System Sciences, Universität Hamburg, Hamburg, Germany, [6]Bjerknes Centre for Climate Research, Geophysical Institute, University of Bergen, Bergen, Norway, [7]Climate and Environmental Physics, Physics Institute, University of Bern, Bern, Switzerland, [8]Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland, [9]European Centre for Medium-Range Weather Forecasts, Reading, UK, [10]Helmholtz Centre for Polar and Marine Research, Alfred Wegener Institute, Bremerhaven, Germany

Abstract The Arctic winter atmospheric boundary layer often features strong and persistent low-level stability (LLS), which arises from longwave radiative cooling of the surface during the polar night. This stable stratification results in a positive lapse rate feedback, which is a major contributor to Arctic amplification. A second state, with cloudy conditions, weaker stability, and near-zero net surface longwave flux is also observed. Previous work has shown that many CMIP5 models fail to appropriately partition water between liquid and ice phases in mixed-phase clouds, leading to a lack of this cloudy state. In this study, we assess the representation of the Arctic winter atmospheric boundary layer over sea ice in global climate models contributing to the latest phase of the Coupled Model Intercomparison Project (CMIP6). We compare boundary layer process relationships in these models to those in surface-based and radiosonde observations collected during the MOSAiC (2019–2020) and SHEBA (1997–1998) expeditions, and by North Pole drifting stations (1955–1991). The majority of CMIP6 models fail to realistically represent the cloudy state over winter Arctic sea ice. Despite this, CMIP6 multimodel mean LLS falls within the observational range, and models mostly capture the observed dependence of LLS on near-surface air temperature and wind speed. CMIP6 models predict a decline in winter LLS with Arctic warming, with mean stability falling below zero by 2100 under the SSP2-4.5 scenario. Our results highlight the failure to accurately simulate mixed-phase clouds as an important limitation on representing a realistic Arctic winter boundary layer in many CMIP6 models.

Plain Language Summary The atmospheric boundary layer is the lowest part of the atmosphere, directly influenced by contact with the Earth's surface. In the Arctic winter, this layer of the atmosphere is often coldest nearest the surface, making it stable against vertical mixing. This feature of Arctic climate is part of the explanation for the much more rapid warming seen in the Arctic than elsewhere, known as Arctic amplification. In this study, we compare the winter Arctic boundary in climate models against that which has been observed in field campaigns. We assess the latest generation of climate models and the most recent major overwinter Arctic field campaign, MOSAiC, as well as earlier and often underexploited observations: the SHEBA campaign and the North Pole drifting stations. We show that cloudy conditions are underrepresented in many of these models. However, these models mostly succeed in representing how stability varies with temperature and wind speed. As the Arctic warms, low-level stability is expected to decrease and models project that the stable state will no longer be dominant in the Arctic winter before the end of the century under a medium emissions scenario.

1. Introduction

The Arctic climate system is undergoing extreme and rapid change. The region is warming roughly four times faster than the global average (Chylek et al., 2022; England et al., 2021; Rantanen et al., 2022), a phenomenon known as Arctic amplification (Previdi et al., 2021; Serreze & Francis, 2006; Serreze et al., 2009; Taylor et al., 2022) and the late-summer sea ice extent has declined by 50% in 40 years (Fetterer et al., 2017, updated 2023). Arctic warming has global consequences; the additional warming due to Arctic amplification is responsible for a time difference of 5 years in the expected crossing date for the Paris agreement's 1.5°C threshold (Duffey et al., 2023). Other consequences include increased global warming due to the albedo feedback

Supervision: Jonathan Day, Felix Pithan
Visualization: Robbie Mallett
Writing – original draft: Alistair Duffey,
Victoria R. Dutch, Julia Steckling
Writing – review & editing:
Alistair Duffey, Robbie Mallett, Victoria
R. Dutch, Antoine Hermant, Jonathan Day,
Felix Pithan

(Pistone et al., 2014), sea level rise due to mass loss from the Greenland ice sheet (Pattyn et al., 2018), an increased greenhouse gas burden from decomposing organic matter in thawed permafrost (Comyn-Platt et al., 2018), and potentially, but controversially (Blackport & Screen, 2020; Cohen et al., 2020), an increased frequency of extreme weather in the Northern Hemisphere midlatitudes (Cohen et al., 2014; Horton et al., 2015; Zhang & Screen, 2021; Zhang et al., 2022).

Arctic amplification is driven by a combination of local feedbacks and remote drivers, including the weaker (negative) Arctic Planck feedback, increasing solar energy absorption with the decline of reflective snow and ice, atmospheric heat transport (e.g., Cai et al., 2022), and the latitudinal variation in deviation from vertically uniform warming (Goosse et al., 2018; Henry et al., 2021; Pithan & Mauritsen, 2014; Taylor et al., 2022). This final contribution, variation in the lapse rate feedback, arises because Arctic warming is more strongly confined near the surface than warming in lower latitudes, and is one of the largest contributors to Arctic amplification in climate models (Goosse et al., 2018; Hahn et al., 2021; Lu & Cai, 2010).

The amplification of Arctic warming is strongest during late autumn and winter (Rantanen et al., 2022; Taylor et al., 2022). During the winter, the atmosphere is often characterized by strong and long-lived low-level stability (LLS), which arises from longwave radiative cooling of the surface leading to atmospheric temperature inversions (Boeke et al., 2021; Wexler, 1936). This stable stratification of the atmospheric boundary layer causes heating to be concentrated near the surface as the Arctic climate warms, resulting in a greater temperature increase at the surface than in the upper troposphere (Manabe & Wetherald, 1975), and a positive local lapse rate feedback (Bintanja et al., 2011; Boeke et al., 2021). Sea ice retreat and atmospheric circulation contribute to this surface-dominant warming profile (Feldl et al., 2020). The vertical profile of Arctic warming contrasts with that seen in the tropics, where convection restores the profile close to the moist adiabat, meaning warming peaks in the upper troposphere and the lapse rate feedback is negative (Hansen et al., 1997).

Given the importance of Arctic wintertime LLS as a control on the lapse rate feedback, accurate simulation of LLS and the processes that control it are vital for reliable predictions of future Arctic climate change in global climate models. Despite this, accurate simulation has been hampered by persistent model biases (Inoue et al., 2021; Pithan et al., 2016; Sedlar et al., 2020; Solomon et al., 2023; Tjernström et al., 2008). In the CMIP5 ensemble, many models fail to realistically represent phenomena specific to the Arctic atmospheric boundary layer, with biases in the phase partition between liquid water and ice in mixed-phase clouds, which impact the modeled LLS (Pithan et al., 2014). Similar biases have also been documented for numerical weather prediction models (Solomon et al., 2023), regional climate models (Inoue et al., 2021; Sedlar et al., 2020), and in the CMIP3 ensemble (Svensson & Karlsson, 2011). Reducing these biases is an important goal of model development, which would enhance the reliability of model projections of Arctic amplification and its local and global consequences. Since CMIP5, modeling centers have made various improvements to their models, which might be expected to improve the representation of mixed-phase clouds, including increased vertical resolution changed convection and microphysics schemes (e.g., Gettelman et al., 2019). Despite this, the extent of these biases in the latest generation of global coupled climate models, the Coupled Model Intercomparison Project (CMIP6) ensemble, has not yet been assessed.

The Arctic atmospheric boundary layer in winter can generally be characterized as being in one of two states (Persson et al., 2002; Stramler et al., 2011): a "clear" state characterized by clear-sky conditions and strong longwave surface cooling ($\gtrsim 50$ Wm^{-2}) due to a lower atmosphere that is relatively transparent to outgoing radiation; and a "cloudy" state in which the surface net longwave radiation is close to zero due to the presence of opaque low-level mixed-phase clouds (Stramler et al., 2011). These two modes also have distinct vertical temperature structures, with stronger atmospheric inversions occurring in the clear state and weaker inversions found further aloft under the cloudy state (Stramler et al., 2011). This bimodal distribution in LLS is found in the CMIP5 model ensemble (Pithan et al., 2014). However, few CMIP5 models showed a realistic representation of the cloudy state, in part due to the freezing of supercooled water droplets at excessively warm temperatures, preventing the formation of mixed-phase clouds with high emissivity (Pithan et al., 2014). Wind speed and shear are also potential sources of model bias in LLS. Higher near-surface wind speeds are associated with increased near-surface air temperature and a reduction in the strength of near-surface temperature inversions over Arctic sea ice (Chechin et al., 2019; Wiel et al., 2017). This is because increased turbulent heat fluxes under high wind speed and shear conditions prevent the formation of strong stability (Chechin et al., 2023; Wiel et al., 2017).

In this study, we investigate the representation of Arctic winter LLS over sea ice and the processes, which control that stability in the latest generation of global climate models, the CMIP6 model ensemble (Eyring et al., 2016). We compare these model representations against observations from the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) campaign (2019–2020) (Shupe et al., 2022), the Surface Heat Budget of the Arctic Ocean (SHEBA) campaign (1997–1998) (Uttal et al., 2002), and 21 North Pole (NP) drifting stations (1954–1990) (Kahl et al., 1999). We build on previous work, which assessed the CMIP5 ensemble against SHEBA observations (Pithan et al., 2014), but here assess the latest generation of CMIP models, and include the recent MOSAiC observations and the extensive NP data.

We first compare the mean-state LLS in CMIP6 climate models over sea ice in winter (November–March, inclusive) to the observations. For models where the required high temporal resolution data are available, we then compare the distribution of clear and cloudy states as defined by net surface longwave radiation to the observed bimodal distribution. Next, we investigate the representation in models of various processes, which control LLS, including the relationship with near-surface air temperature and the suppression of stability by surface winds. Finally, we explore future projections of Arctic winter LLS, and its contribution to Arctic amplification.
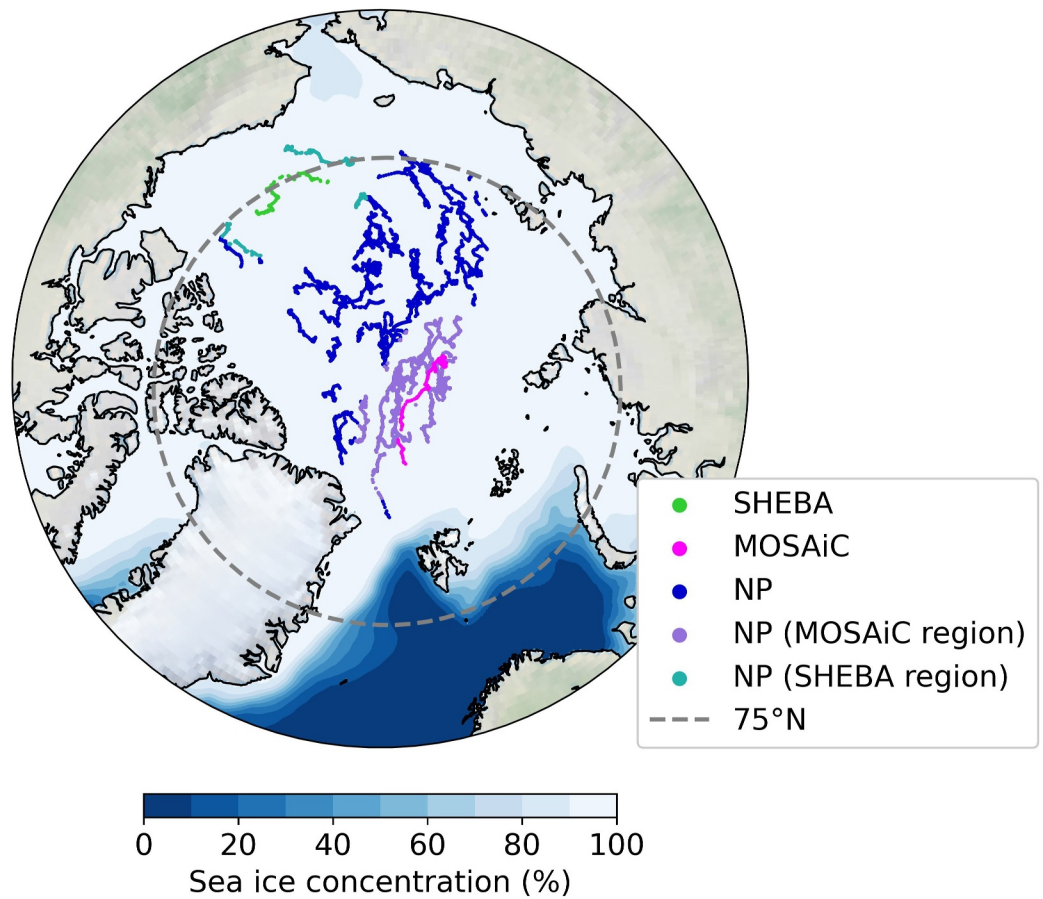
2. Data and Methods

2.1. CMIP6 Data

Our CMIP6 model ensemble is defined by the availability of data at different time resolutions and for different variables on the United Kingdom's CEDA data archive (see Table S1 in Supporting Information S1 for a list of all models used and their DOIs). At monthly resolution, we generally analyze the output of 45 CMIP6 models. However, for the cloud water content and surface heat flux analysis shown in Figure 6, only 37 models have the required data available (see Table S1 in Supporting Information S1). For the analysis of future projections under SSP2-4.5 (Section 3.5), a slightly different set of models was available (see Tables S1 and S2 in Supporting Information S1). For high time resolution data (6-hourly and daily), data availability is more limited and we analyze the outputs of a 20-model subset of the original 45 models. We use the first ensemble member only for each model in constructing our multimodel ensemble. Except when considering future projections (Section 3.5), the time period selected for model output is the final 20 years of the historical scenario (1995–2015) for monthly outputs (Figure 1) and the final 5 years (2010–2015) for daily and subdaily outputs (Figures 2–5). Throughout, we show model output over the extended winter (November–March, inclusive) sea ice domain. This is derived by first regridding each model's monthly sea ice concentration onto that model's atmospheric grid using bilinear interpolation, and then defining a time-varying sea ice mask for atmospheric variables from that model to include data for only those grid points with over 95% sea ice concentration in a given month. Where we examine future trends, we do so for only one emissions scenario, SSP2-4.5 (Eyring et al., 2016), which we select as a medium emissions scenario, which is broadly in line with current global emissions mitigation policies (Hausfather & Peters, 2020).

2.2. Observational Data

We compare model simulations to observational data from three sources (Figure 1). We use radiosonde and surface flux data from MOSAiC (Shupe et al., 2022), which took place from October 2019 to October 2020. Specifically, we use radiosonde data from Maturilli et al. (2021) and tower radiation data from the Atmospheric Radiation Measurement user facility (Reynolds & Riihimaki, 2019). Sondes were generally launched four times per day, but this frequency was increased during exceptional weather events (Peng et al., 2023). We account for this nonuniformity of time sampling in our reported mean values (see Section 2.3). We analyze 340 atmospheric profiles taken during winter. We use radiosonde profiles and surface radiative flux data from SHEBA (Uttal et al., 2002), conducted from October 1997 to October 1998. Detailed information on the SHEBA radiative flux measurements is given by Persson et al. (2002), and on the radiosonde campaign by Beesley et al. (2000) and Bretherton et al. (1999). Additional information is available at the data source (https://atmos.uw.edu/~roode/SHEBA.html, last accessed on January 2024). Radiosondes were generally launched twice daily in winter, resulting in 276 valid winter atmospheric profiles. Additionally, we include radiosonde profiles from the North Pole drifting stations (NP), (Kahl et al., 1999), which operated from 1954 to 1991. While 31 stations existed in the time period, the archive at the US National Snow and Ice Data Centre does not contain usable observations from

Figure 1. Location of winter (November-March) radiosonde observations used in this study, from the MOSAiC (2019–2020) and SHEBA (1997–1998) campaigns, and from the North Pole drifting stations (NP) (1954–1990). The Arctic Ocean shading shows the mean winter sea-ice concentration over the period 1950–2020 (Walsh & Stewart, 2019). The NP (MOSAiC region) and NP (SHEBA region) points, refer, respectively, to subsets of the NP data which lie within 300 km of an observation from the MOSAiC and SHEBA campaigns.
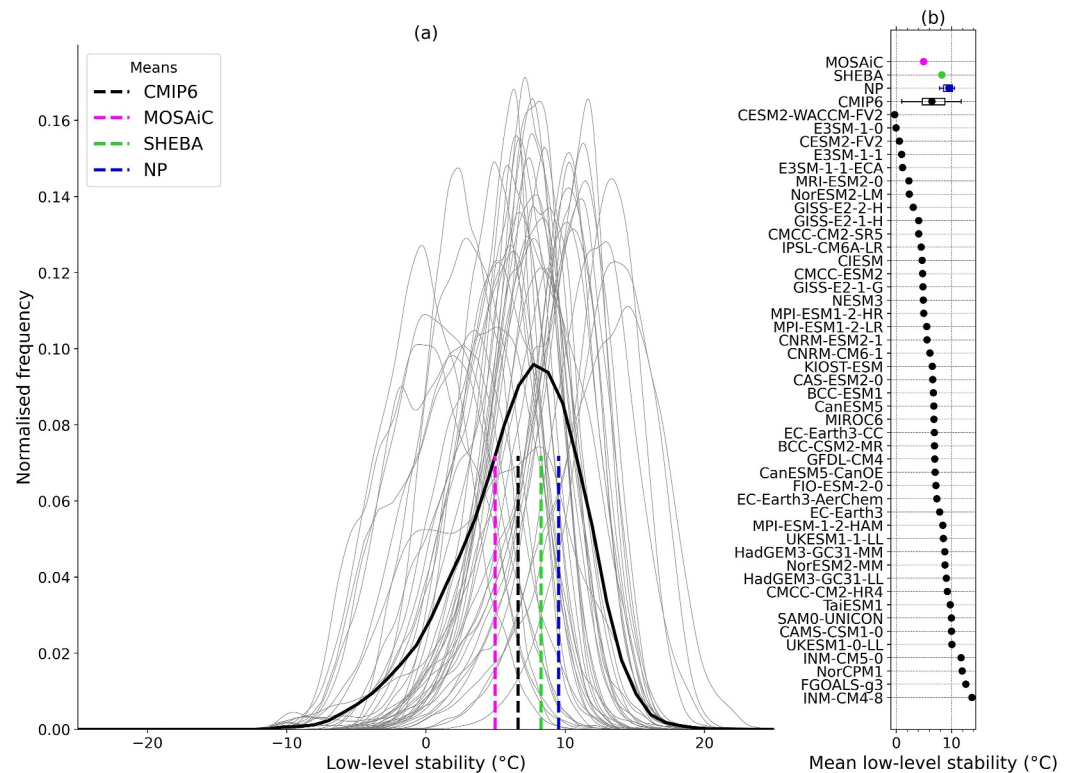
stations 1, 2, 18, 20, 23, 24, 25, 27, 29, and 30. As such, we use data from the remaining 21 stations, which leaves us with 6,999 valid winter atmospheric profiles.

Combined, these three sets of observations include more than 7,500 winter radiosondes, of which we can match 593 (from MOSAiC and SHEBA) to contemporaneous (within 5 min) surface radiative flux measurements (see Section 2.3). These observational campaigns took place under a wide variety of ice conditions, with the icebreaker for MOSAiC surrounded by a loose assemblage of second-year ice (Krumpen et al., 2021) and that of SHEBA surrounded by multiyear ice (Perovich et al., 2003). While sparse in time and space, our observations give coverage across much of the Arctic Ocean domain, over a time period of nearly 70 years. The NP stations also give valuable insight into the interannual variability.

2.3. Methods

Our analysis focuses on the model representation of atmospheric boundary layer processes, which we assess through the relationships between model variables. We assess these relationships over the set of all model data points in time and space during the extended winter period (November- March), hereafter "winter," over sea ice (see Section 2.1). We should not expect a model mean state for any variable over this domain to equal that measured in the observational campaigns. This is because the observations have different regional samplings to the models and because they represent only a single year for SHEBA and MOSAiC, and a set of years with much reduced greenhouse gas forcing for NP. Additionally, the observations are instantaneous and at one point in space, whereas model data are the time step and grid cell mean. As such, we do not focus here on the comparison of

Figure 2. (a) Distribution of winter low-level stability (LLS) over Arctic sea ice (>95% concentration) in CMIP6 models (gray) and field campaigns (colors). LLS is defined as the difference between temperature at 850 hPa and the 2-m air temperature. The bold black line is the multimodel distribution. For all models, the distribution shown is over all grid points in time and space where sea ice concentration is greater than 95%, over the period 1995–2015, and the region north of 65°N. Vertical dashed lines denote mean values. (b) Mean LLS over Arctic sea ice in the winter months (November–March) in CMIP6 models and observations. The "whiskers" on the CMIP6 and NP boxes show the 5th and 95th percentiles of the distribution across models and stations, respectively.

model mean states with observations, but instead on whether the relationships between model variables match those observed in the field. Where the mean state is compared between models and observations (e.g., Figure 2), we account for inconstant temporal sampling over the campaign period in two steps. First, we group the individual sonde observations by date, such that the mean is not biased toward days with additional sonde launches. This is most impactful for MOSAiC, where additional sondes were launched on days of exceptional weather (Peng et al., 2023). Second, we account for inconstant seasonal sampling by first grouping the observational data by month and then taking the full winter mean. Where a "mean" value over time is quoted below for observations, this always refers to a temporally representative value for the campaign's winter season estimated in this manner, as opposed to a simple mean of all observations within the winter period in the data set.

We consider whether each model shows a bimodal distribution in surface net longwave radiation, as a metric for whether that model shows distinct clear and cloudy states (Pithan et al., 2014; Stramler et al., 2011). To derive an objective measure of this bimodality, we use Hartigan's "Dip" test (J. A. Hartigan & Hartigan, 1985; P. M. Hartigan, 1985), which we implement using a Python port of the R package "diptest" (Maechler, 2024). This is a statistical test of whether a unimodal or a multimodal distribution is a better fit to a distribution. We refer to models as "bimodal" in surface net longwave radiation when this test rejects the null hypothesis of unimodality at the 95% significance level.

Following Medeiros et al. (2011) and Pithan et al. (2014), we assess the bulk LLS, defined as the difference between the 850-hPa level temperature and the near-surface (2 m) air temperature, as a proxy for the temperature inversion. LLS is positive when the temperature is warmer aloft. This proxy is useful given the low and variable vertical resolution of CMIP6 models, which complicate direct comparison to radiosonde profiles. Having characterized the bulk LLS following the method above, we also assess its suppression by wind. We do this by

retrieving the surface wind speed of individual sondes in our three observational data sets, and then pairing the speeds with the LLS observed by the same sondes. We produced 189, 341, and 7,753 of these paired data points from the MOSAiC, SHEBA, and North Pole stations, respectively. Surface wind speeds from the North Pole sondes were recorded only to the nearest ms^{-1}. We then compare both the strength and gradient of our observed relationships between LLS and surface wind speed to those in CMIP6 models.

3. Results
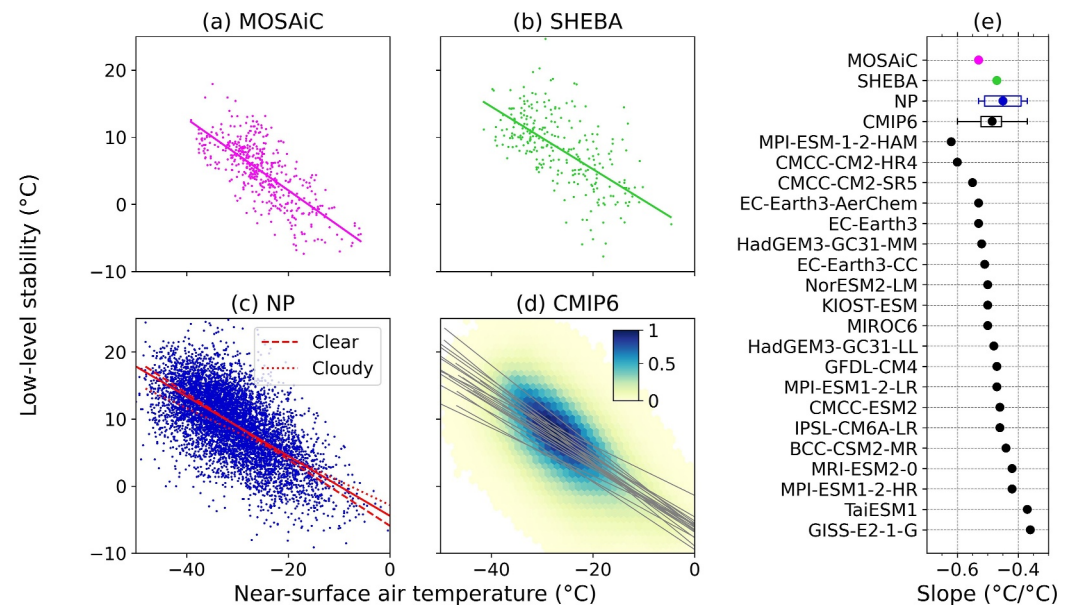
3.1. Mean-State Low-Level Stability

We begin by assessing the LLS over Arctic sea ice during the winter months across the CMIP6 model ensemble and our three observational data sets (Figure 2). The mean LLS values in MOSAiC, SHEBA, and NP are 5.0 ± 0.3, 8.3 ± 0.4, and $9.5 \pm 0.1°C$, respectively, where the errors quoted are the standard error of the mean. We account for uneven temporal sampling in our calculation of these values (see Section 2.3). The three sets of observations were collected at different locations, which will influence their LLS. For example, weaker mean stability is expected in the Atlantic Sector MOSAiC than the Beaufort Gyre SHEBA (Liu et al., 2006). As an approximate means of estimating the impact of this different regional sampling, we also assess subsets of the NP data including only those measurements within 300 km of any observations from first, MOSAiC, and second, SHEBA. The distance threshold of 300 km is selected to sample data from the same Arctic sector while maintaining a reasonable sample size of NP observations within the boundary. In particular, this threshold separates the Pacific and Atlantic sectors, which have different levels of synoptic activity, including the advection of warm and moist air (Woods et al., 2013), and so might differ in the frequency of clear and cloudy states. The locations of these subsets of the NP data are shown in Figure 1. Calculating the mean LLS for these regional subsets of the NP data shows that location is not the primary driver of variation in observed LLS. NP stations within 300 km of the MOSAiC campaign have a mean LLS of $9.3 \pm 0.1°C$, which is lower than the full NP mean, but by only $0.2°C$, as compared to the $4.5°C$ difference between MOSAiC and NP. For NP stations within 300 km of the SHEBA campaign, mean LLS is $9.3 \pm 0.3°C$ which, again, is closer to the full NP mean than the SHEBA mean. As such, the different regional sampling of the campaigns explains only a small fraction of the variation in their observed LLS.

Even accounting for location, we would not expect the mean LLS to be identical across these observational data. Other contributors to the difference include interannual variability—SHEBA and MOSAiC each represent only a single winter—and any long-term temporal trend over the period 1950–2020. The standard deviation of the mean values of LLS for each NP station is $0.94°C$, which gives a rough estimate of the size of interannual variability, and the minimum mean for a station is $7.6°C$. The three sets of observations show a chronological reduction in LLS between the data sources, with the earliest (NP) having the strongest LLS. Because the MOSAiC mean LLS is well outside the range of the NP stations, it seems likely that the decrease in LLS across the three data sets is representative of a wider Arctic trend. Such a negative trend would be expected given the clear negative trend under warming seen in the CMIP6 ensemble (see Section 3.5).

In the CMIP6 multimodel mean, the winter boundary layer over sea ice is characterized by mean LLS of $6.6°C$, as calculated over the final two decades of the historical simulation (1995–2015). This mean value is within the range of mean states ($5–10°C$) seen across the three observational data sets. Two models (CESM2-WACCM-FV2 and E3SM-1-0) out of 45 show negative winter mean LLS over sea ice. The values of mean-state near-surface and 850-hPa temperature for each model, which together define the mean LLS, are shown in Figure S2 in Supporting Information S1. A large majority of models show a strong differentiation between open ocean and sea ice regions, without LLS over the open ocean in the monthly mean state (Figure S1 in Supporting Information S1); only 3 models out of 45 (GISS-E2-1-H, GISS-E2-2-H, and NorCPM1) have positive winter mean-state LLS over the open ocean north of 65°N.

The standard deviation in mean-state LLS over the winter sea ice across the model ensemble is $3.3°C$, and the range between the most and least stable models is $14.0°C$. Pithan et al. (2014) show similar values—a standard deviation of $2.8°C$ and a range of $11.4°C$ (as read from their Figure 4)—for the CMIP5 ensemble, albeit for a slightly different quantity; Pithan et al. take the LLS in the stable mode over the ocean as the "sea-ice" region, and use the months November–February, whereas we here take the "sea-ice" region directly from model sea ice concentrations, for November–March. Similarly, Medeiros et al. (2011) show (as read from their Figure 3) a

Figure 3. Relationship between winter near-surface air temperature and low-level stability, in observations (a–c) and CMIP6 models over sea ice (d). Individual radiosondes are plotted for the observations, and the relative density of time and grid point instances over sea ice as a histogram with normalized density units is plotted for CMIP6 models. Data for all models are 6-hourly, except for NorESM2-LM, IPSL-CM6A-LR, EC-Earth3-CC, and TaiESM1, which use daily mean outputs. Solid lines are linear regressions, with one gray line for each CMIP6 model assessed. In the North Pole drifting stations (NP) case, we also show regressions on the subsets of observations under "clear" and "cloudy" conditions defined as less than and greater than 50% cloud cover by visual assessment, respectively. Panel (e) shows the slopes of the regressions given in (a–d).

standard deviation of 3.1°C and a range of around 10.9°C for the stable-mode LLS across the CMIP3 ensemble. We therefore do not see a substantial reduction in intermodel spread of the mean-state LLS over the winter sea ice between CMIP3, CMIP5, and CMIP6.
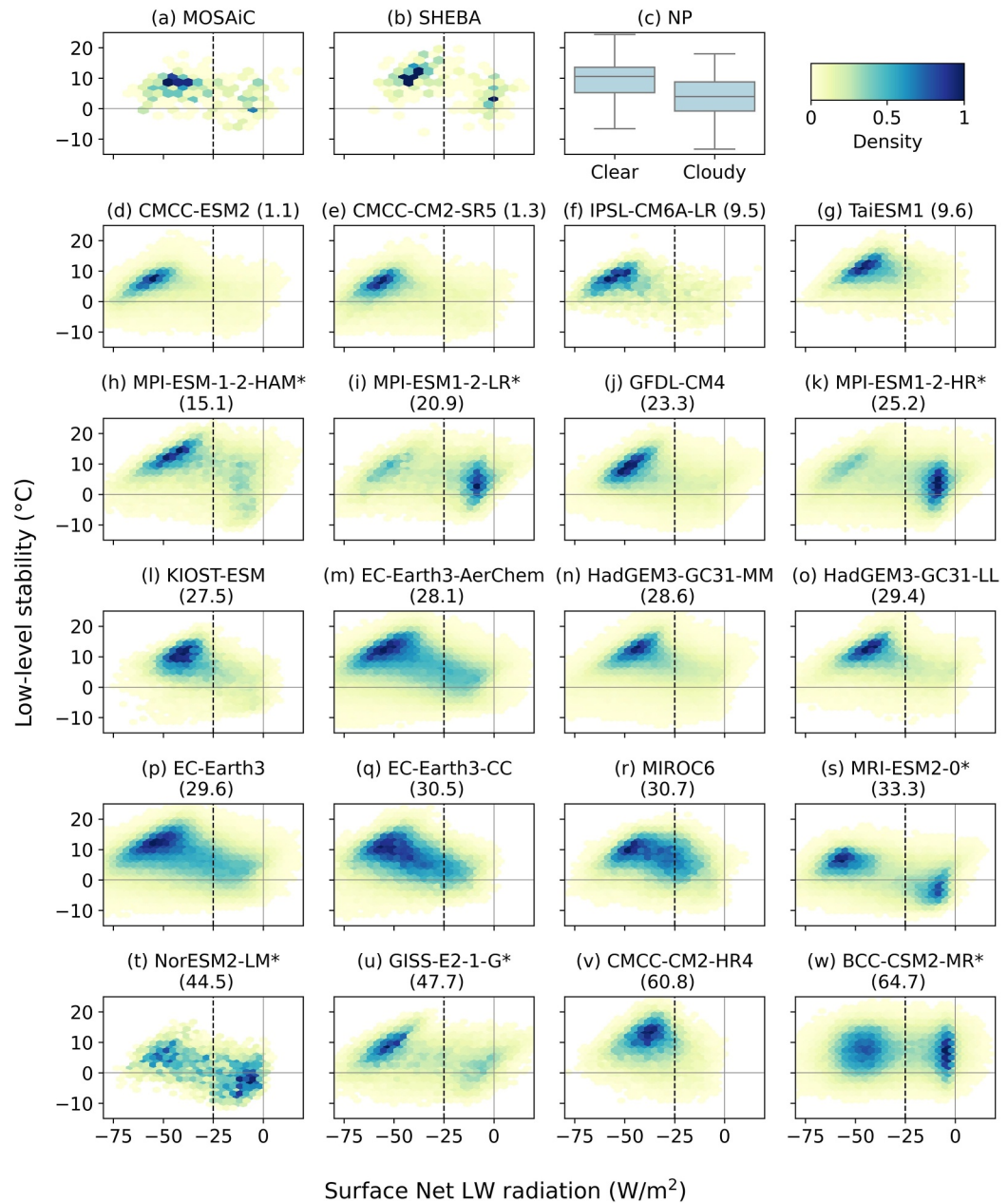
We now investigate how successfully the CMIP6 models represent the processes, which control the winter boundary layer stability over sea ice. We compare the CMIP6 models against observations in terms of their variation in LLS with, first, near-surface air temperature, second, the net surface longwave flux, and third, surface wind strength.

3.2. Near-Surface Air Temperature as a Control on Low-Level Stability

Observations display a linear relationship between LLS and the near-surface air temperature (Figure 3). Such a relationship has been reported previously by Liu et al. (2006), and is consistent between all three observational data sets. LLS decreases by approximately 0.5°C per 1°C increase in near-surface air temperature, that is, the temperature at 850 hPa increases only approximately 0.5°C per 1°C increase at the surface. This relationship is also seen consistently across the ensemble of high time resolution CMIP6 models (see also Figure S3 in Supporting Information S1). The multimodel mean change in LLS with near-surface air temperature is −0.47°C per °C, and the standard deviation across models is 0.06°C per °C. CMIP6 models, therefore, successfully capture the observed coupling on short timescales between temperature at the surface and LLS over sea ice during winter. We also note that, as shown in Figure S2 in Supporting Information S1, a similar linear relationship between near-surface and temperature aloft is also seen across the model ensemble (for each model's mean state), with a model's mean temperature aloft increasing by 0.4°C for each 1°C increase in mean near-surface Arctic winter temperature.

3.3. Clear and Cloudy States

As described in the Introduction, observations support a binary classification of the Arctic winter boundary layer into two states: a clear-sky state with strong longwave radiative cooling of the surface and strong LLS and a cloudy state with near-zero surface net longwave flux and weak or negative LLS (Stramler et al., 2011). By

Figure 4. Bivariate histograms of net longwave radiation at the surface (positive downward) against low-level stability over sea ice, during the winter months (November–March). Histogram units are normalized as the total count of points varies between subplots. Surface radiative flux was not available for the North Pole drifting stations, so NP data are partitioned based on visual assessment of cloud coverage, with coverage of >50% denoted "Cloudy." Data for all models are 6-hourly, except for NorESM2-LM, IPSL-CM6A-LR, EC-Earth3-CC, and TaiESM1, which use daily mean outputs. Vertical black dashed lines show the $-25 \, \mathrm{Wm^{-2}}$ threshold. The values in brackets in each title are the mean-state atmosphere mass content of cloud condensed water (grams $\mathrm{m^{-2}}$) for each model over the same region and the time period; the models are shown in ascending order in this quantity. Starred models are not unimodal in net longwave radiation, according to a dip test at 95% significance (see Section 2.3 and Figure S4 in Supporting Information S1).

plotting bivariate histograms of LLS and net surface longwave (Figure 4), we find distinct clear/cloudy bimodality in the MOSAiC and SHEBA data. We select $-25 \, \mathrm{Wm^{-2}}$ as a threshold, which approximately distinguishes these two modes in both observations and models, as shown in Figure 4. We define any observation with surface net longwave more negative than $-25 \, \mathrm{Wm^{-2}}$ as being in the clear state, which accounts for approximately 60% of observations (MOSAiC 65%, SHEBA 59%, and NP 63%). Surface longwave fluxes time-matched to radiosonde

observations were not available for the NP data set, so we split the LLS observations into two states (Figure 4c), with the clear state defined as less than 50% cloud cover by visual assessment. The proportion of NP observations in the clear state by visual assessment is not strongly sensitive to the choice of threshold because the large majority (79%) of visual observations are either under 10% cloud cover, or over 80% cloud cover. Using alternative clear-sky thresholds of less than 30% or less than 70% would result in the clear state accounting for 57% and 66% of observations, respectively.

Across the CMIP6 models assessed, we find inconsistency in the representation of these two states (clear and cloudy). A visual assessment shows some models, such as MRI-ESM2-0, with a second center of density at close to zero net longwave flux, which we attribute to the cloudy state, while other models, such as TaiESM1, have a single mode at approximately -50 Wm^{-2} net longwave flux, which we identify as having only a clear state. We note that a lack of the cloudy state refers to the lack of high-emissivity, liquid-containing, low clouds, which prevent longwave cooling of the surface. It does not mean that the total cloud fraction (including ice clouds) is zero.
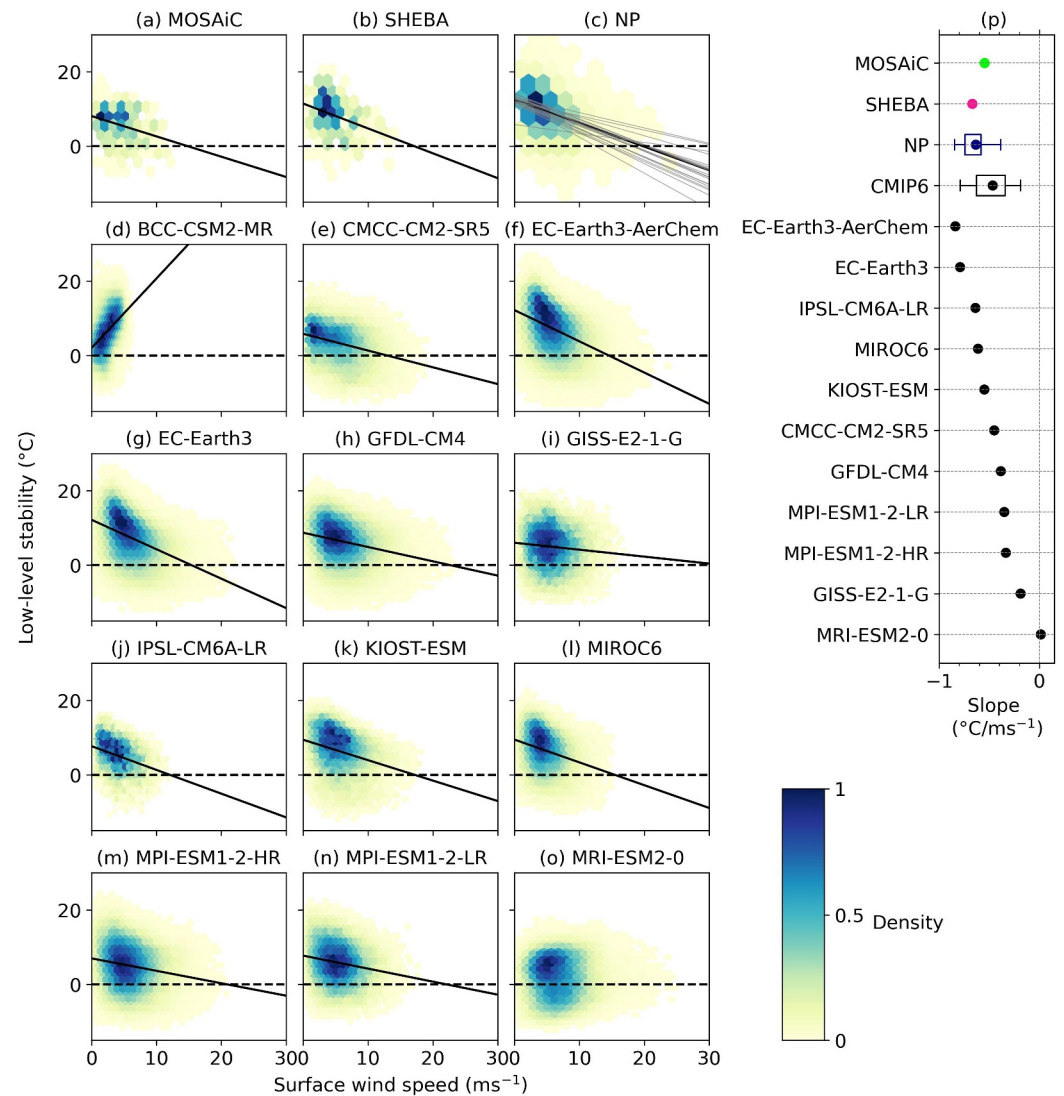
Using a statistical test for multimodality (the dip test, see Section 2.3) on the distribution of 6-hourly net surface longwave flux in each model, we categorize the 20 models in Figure 4 into those which show and those which do not show the two states. The one-dimensional histogram of net surface longwave flux for each model on which this distinction is based is shown in Figure S4 in Supporting Information S1. Seven out of 20 models show a bimodal distribution, with both clear and cloudy atmospheric states (denoted with a * in Figure 4). The ordering of models by their mean column cloud condensed water content (i.e., the ice water path plus the liquid water path) also supports our association of the lack of second mode in LLS with a lack of a cloudy state; those models which lack the cloudy state as assessed based on longwave flux and LLS also show the least atmospheric cloud water content. The spread among models in their cloud condensed water content is very large, ranging between 1 and 65 g m^{-2}. The equivalent mean value for MOSAiC is approximately 80 g m^{-2}, as calculated from the data set of Saavedra Garfias et al. (2023) produced from MOSAiC microwave and radar observations. As such, the higher end of the model range appears more likely to be realistic, although we would expect the MOSAiC mean state to be higher than the regional average shown for the models because moist intrusions are more common in the Atlantic sector of the Arctic (Woods & Caballero, 2016).

For the seven models showing a bimodal LLS distribution, we again use a -25 Wm^{-2} cutoff value to assign data points into one of the two states, as shown in Table S3 in Supporting Information S1. Four models (BCC-CSM2-MR, MPI-ESM1-2-LR, MPI-ESM1-2-HR, and NorESM2-LM) show an over-representation of the cloudy state relative to observations; possible causes of this include cloud microphysics (e.g., excess liquid water) and excess moisture flux into the Arctic (see also Section 3.4). Within the MPI-ESM1 family of models, it appears that model physics rather than resolution is the greater control on realistically representing these states. MPI-ESM-1-2-HAM, which has interactive aerosols as well as altered mixed-phase microphysics, has improved realism in the distribution of clear and cloudy states relative to MPI-ESM1-2-LR and MPI-ESM1-2-HR. Both have a minority of data points in the clear state, while MPI-ESM-1-2-HAM has 60% of points in the clear state, broadly in line with the observations. Alongside the over-representation of the cloudy state, MPI-ESM1-2-LR and MPI-ESM1-2-HR also have mean LLS of 5.7 and 5.0°C, respectively, at the low end of the CMIP6 range, while MPI-ESM-1-2-HAM has stronger stability of 7.7°C.

Comparing the distributions shown in Figure 4 with the mean-state LLS by the model in Figure 2, we observe that the intermodel spread in LLS is determined, not only by the relative frequency of the two modes but also by the central LLS value within each mode. For example, while MRI-ESM2-0 has a realistic 63% of points with under -25 Wm^{-2} net longwave (the cloudy state), it has the lowest mean LLS of the models in Figure 4 because both modes are found at too negative values of LLS.

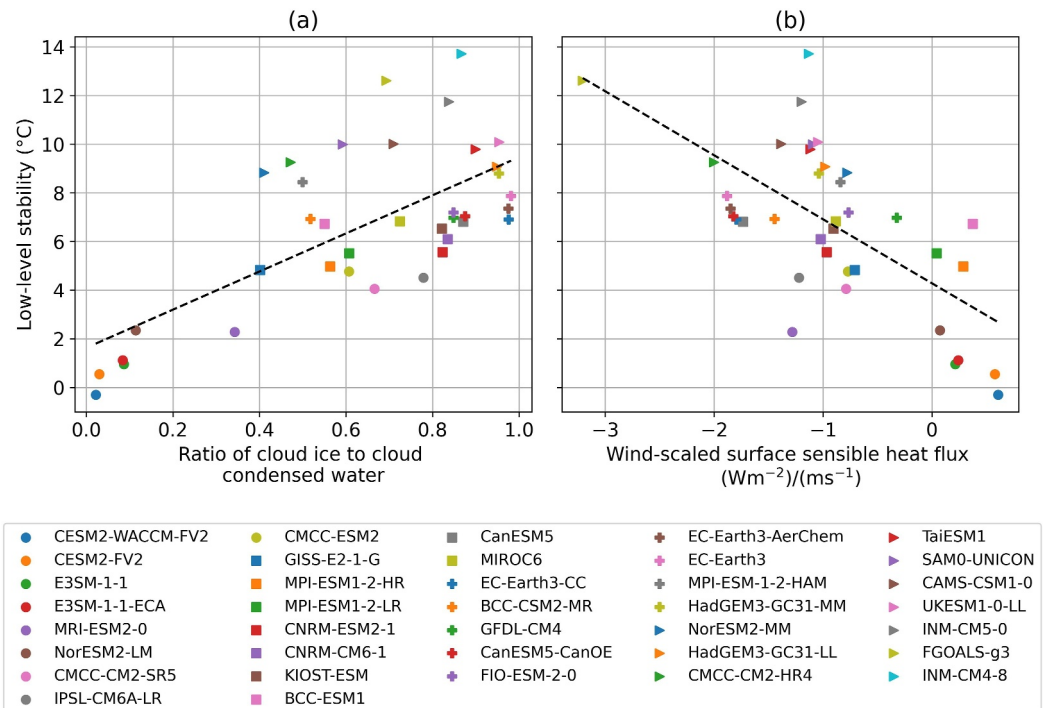### 3.4. Other Controls on Low-Level Stability

We now consider surface wind speed as a control on LLS. The observed relationship between near-surface inversions and wind speed is nonlinear, with several studies finding transition wind speeds, above which the surface inversion strength sharply decreases to near zero (Baas et al., 2019; Vignon et al., 2017; Wiel et al., 2017). At the larger vertical scale of our analysis, such a transition is not apparent (Figures 5a–5c), and we here use a linear regression analysis as a first-order approximation to the nonlinear underlying behavior. In Figure 5, we show this

Figure 5. Relationship between low-level stability and surface wind speed during the winter months (November–March) in each of the three sets of observations (a–c) and the 12 models for which wind data were available (d–o). Solid lines are linear regressions, and colors are bivariate histograms with normalized density units (as the total count of points varies between subplots). For the North Pole drifting stations (NP) (panel c), gray solid lines show the linear regression as calculated for each individual station. All regressions plotted have a p-value less than 0.001. Panel (p) shows the values of the slopes of each linear regression. BCC-CSM2-MR, which has a strongly positive slope, is excluded from (p) and from the CMIP6 distribution shown as a box. For NP, the distribution of slopes across individual stations is shown as a box.

relationship for our three observational data sets, and for 12 CMIP6 models for which wind variables were available at daily (or higher) resolution. The three sets of observations show clear and highly statistical significant reductions in LLS with faster surface wind speed of -0.55, -0.67, and $-0.63°C$ per ms^{-1} in MOSAiC, SHEBA, and NP, respectively, and R^2 values of approximately 0.1 in each case. For NP, this reduction in LLS with increasing wind speed holds both across all stations, and within every individual station.
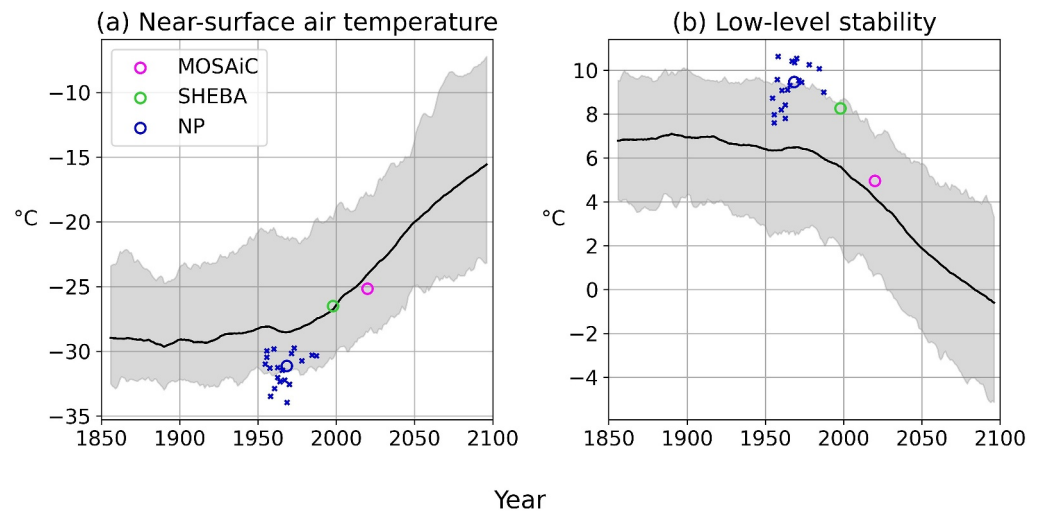
The CMIP6 models show a diverse set of wind speed-LLS relationships. Several models (e.g., IPSL-CM6A-LR) recreate the observed wind suppression closely, with a slope of approximately $-0.6°C$ per ms^{-1} and wind speed accounting for approximately 10% of the variance in LLS. However, one model (BCC-CSM2-MR) shows a strong and significant increase in LLS with wind speed, and another (MRI-ESM2-0) shows almost no relationship, with R^2 less than 0.01. Part of this variation is likely driven by differences in the mean-state wind shear among the models. We would expect an increase in wind-driven atmospheric turbulent heat fluxes with stronger

Figure 6. (a) Ratio of atmospheric column cloud ice to atmospheric column condensed water across the multimodel ensemble in the mean state. (b) Relationship between low-level stability (LLS) and surface sensible heat flux scaled by the surface wind speed. In both panels, each point represents the mean of all monthly instances of LLS and each variable for the first ensemble member over the winter months (November–March) over the sea ice region in a given model over the final two decades of the historical simulation (1995–2015).

wind shear (Chechin et al., 2023), and as a result, a stronger suppression of LLS. This relationship is seen (not shown) in the NP station data, for which there is a positive correlation (significant at 90% confidence) between the mean-state wind shear at a particular station and the reduction in stability with wind speed recorded at that station. The variation in wind shear between NP stations includes contributions from synoptic variability and local sea ice conditions, whereas variability in wind-shear between models, for which the mean state is over the entire winter Arctic sea ice domain over 20 model years in each case, is likely predominantly due to model uncertainty.

Possible contributors to the lack of cloudy state seen in some climate models include synoptic-scale meteorology (e.g., not enough moisture flux into the Arctic), column physics/surface-atmosphere coupling (e.g., bias in vertical turbulent heat fluxes at surface), and cloud microphysics (e.g., early freezing of supercooled water droplets, Pithan et al., 2014). The first of these, synoptic-scale meteorology is out of the scope of this study to assess, but we now make a limited assessment of the potential contribution from the second and third options across the CMIP6 ensemble. Figure 6a shows the relationship between cloud ice fraction and mean-state LLS. The positive relationship here suggests that the freezing of supercooled liquid droplets at excessively warm temperatures and resulting inability to maintain high-emissivity mixed-phase clouds, which contributed to the lack of a cloudy state in CMIP5 models (Pithan et al., 2014), may also persist in CMIP6 models, which lack a cloudy state. However, because we establish only a correlation across the model ensemble here, the reverse relationship is also possible; it may be that models with strong LLS and thus a cold boundary layer see greater condensate freezing. Finally, across the CMIP6 ensemble, there is a strong negative relationship between mean-state surface upward sensible heat flux and LLS over sea ice during the winter months (Figure 6b). This relationship can be explained in terms of the stronger downward fluxes expected given a steeper temperature gradient in the atmospheric column. However, as suggested by Pithan et al. (2014), excessive upward heat fluxes from the warmer underlying ocean due to overestimated ice and snow conductivity may also contribute to reducing LLS in some models.

Figure 7. Time series of the CMIP6 multimodel ensemble projections for (a) high Arctic (>75°N) winter near-surface air temperature change and (b) high Arctic (>75°N) low-level stability. The black line is the multimodel mean of 10-year centered rolling means, and the shaded region is the 10–90th percentile range. The colored dots show the mean values in each observational campaign, with smaller crosses for individual North Pole (NP) stations.

3.5. Projected Decline in Low-Level Stability Under Warming

We have shown (Figure 3) that LLS and near-surface air temperature are negatively correlated in both CMIP6 models and observations on short timescales from hours to days, such that temperatures aloft at 850 hPa increase by ∼0.5°C for each 1°C increase at the surface. Figure S5 in Supporting Information S1 demonstrates that this linear relationship also applies on climatic timescales in each individual CMIP6 model. Over the 250 years of the historical and SSP2-4.5 runs, we find a significant ($p > 99\%$) negative linear relationship between winter near-surface air temperature over Arctic sea ice and LLS in all 40 models assessed. In the multimodel mean, there is a decrease in LLS with near-surface warming of -0.6 ± 0.1°C per °C. This is equivalent to an increase in temperature aloft (850 hPa) of $+0.4$°C per °C near-surface warming. Figure S6 in Supporting Information S1 shows the size of these trends for each model. There is a strong linear relationship between the two trends showing the tight coupling between surface warming, warming aloft (and therefore LLS) across the model ensemble. The 0.4°C warming aloft per °C surface warming of the multimodel mean (i.e., 2.5 times faster warming at the surface than aloft) is a reasonable approximation for the relationship between the trends across the full model ensemble (gray dotted line), although it is not the best linear fit (the black dashed line), which has a nonzero y-intercept of 0.12°C per decade, and slope of 0.33°C per °C.

Every model tested shows faster surface warming than warming at 850 hPa over the 21st century. In each model, therefore, winter LLS declines with the amplified Arctic warming over the 21st century under the SSP2-4.5 emissions scenario. The multimodel mean trajectory is shown in Figure 7. Northward of 75°N, the multimodel mean LLS declines from 6.9°C in the preindustrial period to become negative before the end of the 21st century, with the 10-year rolling mean crossing the zero line in 2083. The rate of decline in LLS across the observations, of -0.85°C per decade, is nearly twice as large as the -0.45°C per decade trend seen in the CMIP6 multimodel mean over the same period. Contributions to this difference could include internal variability and the spatial variation in observation sites, as well as any model limitations.

4. Discussion and Conclusions

In this study, we have assessed the representation of Arctic atmospheric boundary layer processes over sea ice during winter in CMIP6 climate models. Radiosondes from our three data sets find mean wintertime LLS over sea ice of between 5 and 10°C, with the large range likely arising at least in part due to a decreasing trend in LLS over the 70-year time span between the campaigns. The CMIP6 multimodel mean LLS over winter sea ice sits within this range at 6.6°C. However, individual models show a large range of mean states; several models have negative

LLS and thus fail to show a typically stable winter boundary layer, while several other models show a mean stability several degrees stronger than the maximum of the observational range. The intermodel spread in mean-state stability over winter sea ice for CMIP6 does not show a reduction relative to CMIP5 and CMIP3.

Models with high LLS often lack a cloudy state. This lack of a cloudy state, characterized by near-zero net surface longwave flux and weaker, elevated inversions, reported by Pithan et al. (2014) for CMIP5, still applies for a majority of models in this newer CMIP6 generation. As such, improvements in mixed-phase cloud microphysics are still necessary for model development toward accurate simulation of the Arctic winter atmosphere. Despite this, we show that, for the most part, CMIP6 models qualitatively capture the local processes driving variation in LLS. All models assessed reproduce the observed negative linear relationship between near-surface air temperature and LLS in the Arctic winter over sea ice, with approximately a 0.5°C decrease in LLS per °C warming at the surface. Suppression of stability with greater surface winds is also found in the CMIP6 models, albeit less consistently. Both the NP drifting stations and the CMIP6 models show greater wind suppression with increased wind shear. In addition to the negative correlation between near-surface air temperature and stability, which operates on short timescales of hours to days, we also find that a negative linear relationship between near-surface air temperatures and LLS holds on centennial timescales in the CMIP6 models, due to surface-dominant Arctic warming.

Our results do not consider remote drivers of variation in LLS (i.e., moisture and heat fluxes into the Arctic). Future studies might build on our results to include these drivers. In addition, we use only free-running CMIP6 model simulations, and as such, are restricted to a qualitative, process-orientated approach (e.g., Eyring et al., 2005). "Nudged" model runs (e.g., Pithan et al., 2023), which constrain the large-scale atmospheric circulation to a particular state, would allow for a more direct comparison of model outputs against observations on a given day and location. Finally, our assessment of model performance is necessarily limited by the extent of observations; achieving a higher spatial and temporal resolution of observations of the central Arctic during winter would allow for a more quantitative comparison between the modeled and observed states.

As the Arctic warms, the multimodel CMIP6 mean shows winter LLS decreasing to zero in the central Arctic before the end of the century under the medium emissions scenario SSP2-4.5. As such, accurate representation of the less-stable cloudy state may become increasingly important for near-term projections of Arctic climate change using global climate models under increasing warming. Future work might consider in detail these implications. Here, we simply note that this transition away from a typically stably stratified wintertime boundary layer marks yet another profound shift in Arctic climate projected to occur in the coming decades.

Data Availability Statement

All data used in this work are publically available. The CMIP6 data can be accessed from the Earth System Grid Federation CMIP6 archive (https://esgf-index1.ceda.ac.uk/search/cmip6-ceda/). The North Pole drifting stations observations are available to download online from the National Snow and Ice Data Centre, (Colony & Thorndike, 1984). MOSAiC data are available from Pangea (Maturilli et al., 2021) and MOSAiC tower radiation data (Reynolds & Riihimaki, 2019) can be downloaded from https://www.arm.gov. SHEBA data can be downloaded at https://atmos.uw.edu/~roode/SHEBA.html. Code to perform all analysis and plotting is available on Zenodo at https://zenodo.org/records/13763472 (Duffey, 2024).

References

Baas, P., van de Wiel, B. J. H., van Meijgaard, E., Vignon, E., Genthon, C., vander Linden, S. J. A., & de Roode, S. R. (2019). Transitions in the wintertime near-surface temperature inversion at Dome C, Antarctica. *Quarterly Journal of the Royal Meteorological Society*, *145*(720), 930–946. https://doi.org/10.1002/qj.3450

Beesley, J. A., Bretherton, C. S., Jakob, C., Andreas, E. L., Intrieri, J. M., & Uttal, T. A. (2000). A comparison of cloud and boundary layer variables in the ECMWF forecast model with observations at Surface Heat Budget of the Arctic Ocean (SHEBA) ice camp. *Journal of Geophysical Research*, *105*(D10), 12337–12349. https://doi.org/10.1029/2000JD900079

Bintanja, R., Graversen, R. G., & Hazeleger, W. (2011). Arctic winter warming amplified by the thermal inversion and consequent low infrared cooling to space. *Nature Geoscience*, *4*(11), 758–761. https://doi.org/10.1038/ngeo1285

Blackport, R., & Screen, J. A. (2020). Weakened evidence for mid-latitude impacts of Arctic warming. *Nature Climate Change*, *10*(12), 1065–1066. https://doi.org/10.1038/s41558-020-00954-y

Boeke, R. C., Taylor, P. C., & Sejas, S. A. (2021). On the nature of the Arctic's positive lapse-rate feedback. *Geophysical Research Letters*, *48*(1), e2020GL091109. https://doi.org/10.1029/2020GL091109

Bretherton, C. S., de Roode, S. R., Jakob, C., Andreas, E. L., Intrieri, J., & Persson, P. O. G. (1999). A comparison of the ECMWF forecast model with observations over the annual cycle at SHEBA. *Journal of Geophysical Research (FIRE Arctic Clouds Experiment Special Issue)*. Retrieved from https://dspace.library.uu.nl/bitstream/handle/1874/2608/sheba2.pdf?sequence=1&isAllowed=y

Cai, Z., You, Q., Chen, H. W., Zhang, R., Chen, D., Chen, J., et al. (2022). Amplified wintertime Barents Sea warming linked to intensified Barents oscillation. *Environmental Research Letters*, *17*(4), 044068. https://doi.org/10.1088/1748-9326/ac5bb3

Chechin, D. G., Lüpkes, C., Hartmann, J., Ehrlich, A., & Wendisch, M. (2023). Turbulent structure of the Arctic boundary layer in early summer driven by stability, wind shear and cloud-top radiative cooling: ACLOUD airborne observations. *Atmospheric Chemistry and Physics*, *23*(8), 4685–4707. https://doi.org/10.5194/acp-23-4685-2023

Chechin, D. G., Makhotina, I. A., Lüpkes, C., & Makshtas, A. P. (2019). Effect of wind speed and leads on clear-sky cooling over Arctic Sea Ice during polar night. *Journal of the Atmospheric Sciences*, *76*(8), 2481–2503. https://doi.org/10.1175/JAS-D-18-0277.1

Chylek, P., Folland, C., Klett, J. D., Wang, M., Hengartner, N., Lesins, G., & Dubey, M. K. (2022). Annual mean Arctic amplification 1970–2020: Observed and simulated by CMIP6 climate models. *Geophysical Research Letters*, *49*(13), e2022GL099371. https://doi.org/10.1029/2022GL099371

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., et al. (2014). Recent Arctic amplification and extreme mid-latitude weather. *Nature Geoscience*, *7*(9), 627–637. https://doi.org/10.1038/ngeo2234

Cohen, J., Zhang, X., Francis, J., Jung, T., Kwok, R., Overland, J., et al. (2020). Divergent consensuses on Arctic amplification influence on midlatitude severe winter weather. *Nature Climate Change*, *10*(1), 20–29. https://doi.org/10.1038/s41558-019-0662-y

Colony, R., & Thorndike, A. S. (1984). *Arctic Ocean drift tracks from ships, buoys, and manned research stations, 1872–1973, version 1*. National Snow and Ice Data Center. https://doi.org/10.7265/N5D798B1

Comyn-Platt, E., Hayman, G., Huntingford, C., Chadburn, S. E., Burke, E. J., Harper, A. B., et al. (2018). Carbon budgets for 1.5 and 2°C targets lowered by natural wetland and permafrost feedbacks. *Nature Geoscience*, *11*(8), 568–573. https://doi.org/10.1038/s41561-018-0174-9

Duffey, A. (2024). alistairduffey/ABL_cmip6:v1. Zenodo. https://doi.org/10.5281/zenodo.13763472

Duffey, A., Mallett, R., Irvine, P. J., Tsamados, M., & Stroeve, J. (2023). ESD Ideas: Arctic amplification's contribution to breaches of the Paris Agreement. *Earth System Dynamics*, *14*(6), 1165–1169. https://doi.org/10.5194/esd-14-1165-2023

England, M. R., Eisenman, I., Lutsko, N. J., & Wagner, T. J. W. (2021). The recent emergence of Arctic amplification. *Geophysical Research Letters*, *48*(15), e2021GL094086. https://doi.org/10.1029/2021GL094086

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model inter-comparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Eyring, V., Harris, N. R. P., Rex, M., Shepherd, T. G., Fahey, D. W., Amanatidis, G. T., et al. (2005). A strategy for process-oriented validation of coupled chemistry–climate models. *Bulletin of the American Meteorological Society*, *86*(8), 1117–1134. https://doi.org/10.1175/BAMS-86-8-1117

Feldl, N., Po-Chedley, S., Singh, H. K. A., Hay, S., & Kushner, P. J. (2020). Sea ice and atmospheric circulation shape the high-latitude lapse rate feedback. *npj Climate and Atmospheric Science*, *3*(1), 1–9. https://doi.org/10.1038/s41612-020-00146-7

Fetterer, F., Knowles, K., Meier, W. N., Savoie, M., & Windnagel, A. K. (2017). *Sea ice index, version 3*. NSIDC: National Snow and Ice Data Center. https://doi.org/10.7265/N5K072F8

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., et al. (2019). High climate sensitivity in the community Earth system model version 2 (CESM2). *Geophysical Research Letters*, *46*(14), 8329–8337. https://doi.org/10.1029/2019GL083978

Goosse, H., Kay, J. E., Armour, K. C., Bodas-Salcedo, A., Chepfer, H., Docquier, D., et al. (2018). Quantifying climate feedbacks in polar regions. *Nature Communications*, *9*(1), 1919. https://doi.org/10.1038/s41467-018-04173-0

Hahn, L. C., Armour, K. C., Zelinka, M. D., Bitz, C. M., & Donohoe, A. (2021). Contributions to polar amplification in CMIP5 and CMIP6 models. *Frontiers in Earth Science*, *9*, 710036. https://doi.org/10.3389/feart.2021.710036

Hansen, J., Sato, M., & Ruedy, R. (1997). Radiative forcing and climate response. *Journal of Geophysical Research*, *102*(D6), 6831–6864. https://doi.org/10.1029/96JD03436

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, *13*(1), 70–84. https://doi.org/10.1214/aos/1176346577

Hartigan, P. M. (1985). Computation of the dip statistic to test for unimodality. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *34*(3), 320–325. https://doi.org/10.2307/2347485

Hausfather, Z., & Peters, G. P. (2020). RCP8.5 is a problematic scenario for near-term emissions. *Proceedings of the National Academy of Sciences*, *117*(45), 27791–27792. https://doi.org/10.1073/pnas.2017124117

Henry, M., Merlis, T. M., Lutsko, N. J., & Rose, B. E. J. (2021). Decomposing the drivers of polar amplification with a single-column model. *Journal of Climate*, *34*(6), 2355–2365. https://doi.org/10.1175/JCLI-D-20-0178.1

Horton, D. E., Johnson, N. C., Singh, D., Swain, D. L., Rajaratnam, B., & Diffenbaugh, N. S. (2015). Contribution of changes in atmospheric circulation patterns to extreme temperature trends. *Nature*, *522*(7557), 465–469. https://doi.org/10.1038/nature14550

Inoue, J., Sato, K., Rinke, A., Cassano, J. J., Fettweis, X., Heinemann, G., et al. (2021). Clouds and radiation processes in regional climate models evaluated using observations over the ice-free Arctic Ocean. *Journal of Geophysical Research: Atmospheres*, *126*(1), e2020JD033904. https://doi.org/10.1029/2020JD033904

Kahl, J. D. W., Zaitseva, N. A., Khattatov, V., Schnell, R. C., Bacon, D. M., Bacon, J., et al. (1999). Radiosonde observations from the former Soviet "North Pole" series of drifting ice stations, 1954–90. *Bulletin of the American Meteorological Society*, *80*(10), 2019–2026. https://doi.org/10.1175/1520-0477(1999)080<2019:roftfs>2.0.co;2

Krumpen, T., von Albedyll, L., Goessling, H. F., Hendricks, S., Juhls, B., Spreen, G., et al. (2021). MOSAiC drift expedition from October 2019 to July 2020: Sea ice conditions from space and comparison with previous years. *The Cryosphere*, *15*(8), 3897–3920. https://doi.org/10.5194/tc-15-3897-2021

Liu, Y., Key, J. R., Schweiger, A., & Francis, J. (2006). Characteristics of satellite-derived clear-sky atmospheric temperature inversion strength in the Arctic, 1980–96. *Journal of Climate*, *19*(19), 4902–4913. https://doi.org/10.1175/JCLI3915.1

Lu, J., & Cai, M. (2010). Quantifying contributions to polar warming amplification in an idealized coupled general circulation model. *Climate Dynamics*, *34*(5), 669–687. https://doi.org/10.1007/s00382-009-0673-x

Maechler, M. (2024). diptest: Hartigan's Dip test statistic for unimodality. Retrieved from https://github.com/mmaechler/diptest

Manabe, S., & Wetherald, R. T. (1975). The effects of doubling the CO_2 concentration on the climate of a general circulation model. *Journal of the Atmospheric Sciences*, *32*(1), 3–15. https://doi.org/10.1175/1520-0469(1975)032⟨0003:TEODTC⟩2.0.CO;2

Maturilli, M., Holdridge, D. J., Dahlke, S., Graeser, J., Sommerfeld, A., Jaiser, R., et al. (2021). *Initial radiosonde data from 2019-10 to 2020-09 during project MOSAiC*. PANGAEA. https://doi.org/10.1594/PANGAEA.928656

Medeiros, B., Deser, C., Tomas, R. A., & Kay, J. E. (2011). Arctic inversion strength in climate models. *Journal of Climate*, *24*(17), 4733–4740. https://doi.org/10.1175/2011JCLI3968.1

Pattyn, F., Ritz, C., Hanna, E., Asay-Davis, X., DeConto, R., Durand, G., et al. (2018). The Greenland and Antarctic ice sheets under 1.5°C global warming. *Nature Climate Change*, *8*(12), 1053–1061. https://doi.org/10.1038/s41558-018-0305-8

Peng, S., Yang, Q., Shupe, M. D., Xi, X., Han, B., Chen, D., et al. (2023). The characteristics of atmospheric boundary layer height over the Arctic Ocean during MOSAiC. *Atmospheric Chemistry and Physics*, *23*(15), 8683–8703. https://doi.org/10.5194/acp-23-8683-2023

Perovich, D. K., Grenfell, T. C., Richter-Menge, J. A., Light, B., Tucker, III, W. B., & Eicken, H. (2003). Thin and thinner: Sea ice mass balance measurements during SHEBA. *Journal of Geophysical Research*, *108*(C3). 8050. https://doi.org/10.1029/2001JC001079

Persson, P. O. G., Fairall, C. W., Andreas, E. L., Guest, P. S., & Perovich, D. K. (2002). Measurements near the atmospheric surface flux Group tower at SHEBA: Near-surface conditions and surface energy budget. *Journal of Geophysical Research*, *107*(C10), 21. https://doi.org/10.1029/2000JC000705

Pistone, K., Eisenman, I., & Ramanathan, V. (2014). Observational determination of albedo decrease caused by vanishing Arctic sea ice. *Proceedings of the National Academy of Sciences*, *111*(9), 3322–3326. https://doi.org/10.1073/pnas.1318201111

Pithan, F., Ackerman, A., Angevine, W. M., Hartung, K., Ickes, L., Kelley, M., et al. (2016). Select strengths and biases of models in representing the Arctic winter boundary layer over sea ice: The larcform 1 single column model intercomparison. *Journal of Advances in Modeling Earth Systems*, *8*(3), 1345–1357. https://doi.org/10.1002/2016MS000630

Pithan, F., Athanase, M., Dahlke, S., Sánchez-Benítez, A., Shupe, M. D., Sledd, A., et al. (2023). Nudging allows direct evaluation of coupled climate models with in situ observations: A case study from the MOSAiC expedition. *Geoscientific Model Development*, *16*(7), 1857–1873. https://doi.org/10.5194/gmd-16-1857-2023

Pithan, F., & Mauritsen, T. (2014). Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience*, *7*(3), 181–184. https://doi.org/10.1038/ngeo2071

Pithan, F., Medeiros, B., & Mauritsen, T. (2014). Mixed-phase clouds cause climate model biases in Arctic wintertime temperature inversions. *Climate Dynamics*, *43*(1), 289–303. https://doi.org/10.1007/s00382-013-1964-9

Previdi, M., Smith, K. L., & Polvani, L. M. (2021). Arctic amplification of climate change: A review of underlying mechanisms. *Environmental Research Letters*, *16*(9), 093003. https://doi.org/10.1088/1748-9326/ac1c29

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., et al. (2022). The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth and Environment*, *3*(1), 1–10. https://doi.org/10.1038/s43247-022-00498-3

Reynolds, R., & Riihimaki, L. (2019). Arm: Icerad. https://doi.org/10.5439/1608608

Saavedra Garfias, P., Kalesse-Los, H., von Albedyll, L., Griesche, H., & Spreen, G. (2023). Asymmetries in cloud microphysical properties ascribed to sea ice leads via water vapour transport in the central Arctic. *Atmospheric Chemistry and Physics*, *23*(22), 14521–14546. https://doi.org/10.5194/acp-23-14521-2023

Sedlar, J., Tjernström, M., Rinke, A., Orr, A., Cassano, J., Fettweis, X., et al. (2020). Confronting Arctic troposphere, clouds, and surface energy budget representations in regional climate models with observations. *Journal of Geophysical Research: Atmospheres*, *125*(6), e2019JD031783. https://doi.org/10.1029/2019JD031783

Serreze, M. C., Barrett, A. P., Stroeve, J. C., Kindig, D. N., & Holland, M. M. (2009). The emergence of surface-based Arctic amplification. *The Cryosphere*, *3*(1), 11–19. https://doi.org/10.5194/tc-3-11-2009

Serreze, M. C., & Francis, J. A. (2006). The Arctic amplification debate. *Climatic Change*, *76*(3), 241–264. https://doi.org/10.1007/s10584-005-9017-y

Shupe, M. D., Rex, M., Blomquist, B., Persson, P. O. G., Schmale, J., Uttal, T., et al. (2022). Overview of the MOSAiC expedition: Atmosphere. *Elementa: Science of the Anthropocene*, *10*(1), 00060. https://doi.org/10.1525/elementa.2021.00060

Solomon, A., Shupe, M. D., Svensson, G., Barton, N. P., Batrak, Y., Bazile, E., et al. (2023). The winter central Arctic surface energy budget: A model evaluation using observations from the MOSAiC campaign. *Elementa: Science of the Anthropocene*, *11*(1), 00104. https://doi.org/10.1525/elementa.2022.00104

Stramler, K., Genio, A. D. D., & Rossow, W. B. (2011). Synoptically driven Arctic winter states. *Journal of Climate*, *24*(6), 1747–1762. https://doi.org/10.1175/2010JCLI3817.1

Svensson, G., & Karlsson, J. (2011). On the Arctic wintertime climate in global climate models. *Journal of Climate*, *24*(22), 5757–5771. https://doi.org/10.1175/2011JCLI4012.1

Taylor, P. C., Boeke, R. C., Boisvert, L. N., Feldl, N., Henry, M., Huang, Y., et al. (2022). Process drivers, inter-model spread, and the path forward: A review of amplified Arctic warming. *Frontiers in Earth Science*, *9*, 758361. https://doi.org/10.3389/feart.2021.758361

Tjernström, M., Sedlar, J., & Shupe, M. D. (2008). How well do regional climate models reproduce radiation and clouds in the Arctic? An evaluation of ARCMIP simulations. *Journal of Applied Meteorology and Climatology*, *24*(22), 5757–5771. https://doi.org/10.1175/2008JAMC1845.1

Uttal, T., Curry, J. A., McPhee, M. G., Perovich, D. K., Moritz, R. E., Maslanik, J. A., et al. (2002). Surface heat budget of the Arctic Ocean. *Bulletin of the American Meteorological Society*, *83*(2), 255–276. https://doi.org/10.1175/1520-0477(2002)083⟨0255:SHBOTA⟩2.3.CO;2

Vignon, E., van de Wiel, B. J. H., van Hooijdonk, I. G. S., Genthon, C., vander Linden, S. J. A., van Hooft, J. A., et al. (2017). Stable boundary-layer regimes at dome C, Antarctica: Observation and analysis. *Quarterly Journal of the Royal Meteorological Society*, *143*(704), 1241–1253. https://doi.org/10.1002/qj.2998

Walsh, W. L. C. F. F. J. E., & Stewart, J. S. (2019). *Gridded monthly sea ice extent and concentration, 1850 onward*. Version 2. National Snow and Ice Data Center. https://doi.org/10.7265/jj4s-tq79

Wexler, H. (1936). Cooling in the lower atmosphere and the structure of polar continental air. *Monthly Weather Review*, *64*(4), 122–136. https://doi.org/10.1175/1520-0493(1936)64⟨122:CITLAA⟩2.0.CO;2

Wiel, B. J. H. V. D., Vignon, E., Baas, P., Hooijdonk, I. G. S. V., Linden, S. J. A. V. d., Hooft, J. A. V., et al. (2017). Regime transitions in near-surface temperature inversions: A conceptual model. *Journal of the Atmospheric Sciences*, *74*(4), 1057–1073. https://doi.org/10.1175/JAS-D-16-0180.1

Woods, C., & Caballero, R. (2016). The role of moist intrusions in winter Arctic warming and Sea Ice decline. *Journal of Climate*, *29*(12), 4473–4485. https://doi.org/10.1175/JCLI-D-15-0773.1

Woods, C., Caballero, R., & Svensson, G. (2013). Large-scale circulation associated with moisture intrusions into the Arctic during winter. *Geophysical Research Letters*, *40*(17), 4717–4721. https://doi.org/10.1002/grl.50912

Zhang, R., & Screen, J. A. (2021). Diverse Eurasian winter temperature responses to barents-Kara Sea Ice anomalies of different magnitudes and seasonality. *Geophysical Research Letters*, *48*(13), e2021GL092726. https://doi.org/10.1029/2021GL092726

Zhang, R., Screen, J. A., & Zhang, R. (2022). Arctic and Pacific Ocean conditions were favorable for cold extremes over Eurasia and North America during winter 2020/21. *Bulletin of the American Meteorological Society*, *103*(10), E2285–E2301. https://doi.org/10.1175/BAMS-D-21-0264.1