Too cold, too saturated? Evaluating climate models at the gateway to the Arctic

Felix Pithan[1], Ann Kristin Naumann[2,3,4], and Marion Maturilli[1]

[1]Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Potsdam/Bremerhaven, Germany
[2]Max Planck Institute for Meteorology, Hamburg, Germany
[3]Meteorological Institute, Universität Hamburg, Hamburg, Germany
[4]now at Ludwig-Maximilians-Universität, Munich, Germany

Correspondence: Felix Pithan (felix.pithan@awi.de)

Abstract. The Arctic wintertime energy and moisture ~~budget~~ budgets are largely controlled by the advection of warm, moist air masses from lower latitudes, cooling and drying of these air masses inside the Arctic and the export of cold, dry air masses. Climate models have substantial difficulties in representing key processes in these air-mass transformations, including turbulence under stable stratification and mixed-phase cloud processes. Here, we use radiosonde profiles of temperature and moisture and surface radiation observations from Ny Ålesund, Svalbard (1993-2014), to assess the properties of air masses being imported into and exported from the central Arctic in CMIP6 climate models. In the free troposphere, models tend to be cold-biased especially for the coldest temperatures, and relative humidity in most models is closer to saturation with respect to ice than observed. In the analysed models, supersaturation with respect to ice tends to be better represented with two-moment microphysics. The overall distribution of column-integrated precipitable water in models matches well with observations. Cold and dry biases are stronger in air masses being exported from the Arctic than those entering the Arctic. This suggests that the previously reported cold bias in the Arctic in CMIP6 models is probably due to errors in local thermodynamic processes.

1 Introduction

The Arctic radiates more energy to space than it receives from the sun, and the resulting energy deficit is compensated by heat convergence in the atmosphere and ocean, especially in winter (Mayer et al., 2019). The atmospheric energy and moisture convergence largely occurs through the exchange of air masses between the Arctic and lower latitudes (Pithan et al., 2018). Warm, moist air masses are advected polewards, where they cool and dry (Wexler, 1936; Curry, 1983; Ali and Pithan, 2020), whereas cold, dry air masses leave the Arctic and pick up heat and moisture over the open ocean in marine cold-air outbreaks (Papritz and Sodemann, 2018).

Arctic air-mass transformations are driven by and connected to the Arctic surface energy budget by thermodynamic processes in clouds and boundary layers. ~~These processes are~~ It is challenging to represent these processes in climate models, and ~~previous~~ earlier generations of climate models ~~had~~ have substantial biases in their representation of Arctic climate, including the vertical temperature structure (Svensson and Karlsson, 2011; Medeiros et al., 2011; Pithan et al., 2014). Some of these biases, such as a pronounced cold bias over the central Arctic ocean in winter, persist in CMIP6 models (Davy and Outten, 2020). The

convergence of heat and moisture in the Arctic atmosphere depends on the amount of air exchanged between the Arctic and lower latitudes, which is controlled by the large-scale circulation (Woods et al., 2013; Graversen and Burtu, 2016), and on the properties, especially the heat and moisture content, of the transported air.

The particular vertical temperature structure of the Arctic atmosphere with frequent stable stratification ~~that is shaped by Arctic air-mass transformations~~ is mainly due to atmospheric heat advection from the south and diabatic cooling of the surface, and is an important condition for Arctic amplification of climate change (Manabe and Wetherald, 1975; Pithan and Mauritsen, 2014; Boeke et al., 2021).

Clouds play an important role in Arctic air-mass transformations and for the Arctic surface energy budget (Curry, 1983; Karlsson and Svensson, 2011; Cronin and Tziperman, 2015). Throughout most of the year, the warming longwave radiative effect of clouds dominates over the cooling shortwave radiative effect in the Arctic, i.e. clouds have a net warming effect on the surface (Intrieri et al., 2002). Clouds increase the emissivity of the atmosphere and can thereby increase atmospheric radiative cooling and precipitation formation (Pithan and Jung, 2021; Bonan et al., 2024).

In general, clouds form when an air parcel is (super)saturated with water vapour, i.e. when it contains more water vapour than it can sustain in vapourised form, such that the excess water vapour condenses into cloud droplets or freezes. Because the moisture content of the atmosphere fluctuates at relatively small horizontal scales (Quaas, 2012), cloud formation can occur on much smaller scales than resolved by climate models with grid spacings of tens of kilometers, and hence needs to be parameterized in such models.

Accurate representation of cloud phase is difficult in climate models, but can have an important impact on cloud feedbacks and thus future climate change (Cesana et al., 2022; Tan et al., 2022). At the same temperature and pressure, the saturation vapour pressure over a liquid water surface is higher than that over an ice surface~~(Wegener, 1911; Bergeron, 1928; Findeisen, 1938; Storelv~~. Air can thus be supersaturated with respect to ice, but still subsaturated with respect to liquid water. In this regime, water droplets can quickly evaporate, while ice crystals grow through deposition of water molecules from the gas phase (Wegener, 1911; Bergeron .

In real-world mixed-phase clouds, supercooled liquid water can be concentrated in thin (~~O 10m~~few tens of meters) layers near cloud top (Verlinde et al., 2007), whereas models usually assume all condensate to be homogeneously distributed within the cloudy part of a grid box. At the same time, climate models differ in the degree of complexity used to represent ~~mixed phase~~ mixed-phase cloud microphysics, and the respective roles of resolution and realism of microphysical process representations for model biases are unclear.

Observational records of Arctic climate are scarce, and existing climate model evaluations largely focus on surface fields and large-scale climatological means (Davy and Outten, 2020), or remain qualitative in comparing model output to shorter field campaigns (Pithan et al., 2014; Linke et al., 2023). In this paper, we evaluate the properties of air masses exchanged between the Arctic and lower latitudes in CMIP6 models using radiosonde and surface radiation observations from Ny Ålesund, Svalbard (Maturilli et al., 2015; Maturilli and Kayser, 2017a), which is located within the major pathway for moist intrusions entering the Arctic in winter (Woods and Caballero, 2016). We stratify observations and sub-daily model output by wind directions to assess poleward-moving and equatorward-moving air masses separately and investigate the distribution of relative humidity.

We test the effect of horizontal resolution vs. parametrization complexity on the representation of supersaturation with respect to ice using output from a km-scale global model and a sensitivity experiment of prescribed (1-moment microphysics) vs. prognostic (2-moment microphysics) ice number concentrations.

2 Data and methods

2.1 Observations

We use wintertime (December to March) meteorological upper-air observations by radiosondes from Ny-Ålesund, Svalbard, manually launched by the Alfred Wegener Institute since 1993. During these three decades, different types of radiosondes were used for the soundings, potentially introducing instrument related inhomogeneities to the data record. We therefore rely on the homogenized radiosonde dataset for Ny-Ålesund (Maturilli and Kayser, 2016; Maturilli and Kayser, 2017b; Maturilli and Dünschede, 2023), which corrects known biases of each radiosonde type (Maturilli and Kayser, 2017a). The sounding data are interpolated to the vertical resolution of the CMIP6 model output (the plev 27 levels as defined in Juckes et al. (2020), see their table 4). For comparison with CMIP6 models, we use the period of overlap with the historical runs, i.e. 1993-2014. To identify cloudy and clear sky conditions, we furthermore use longwave radiation measurements from Ny-Ålesund (Maturilli et al., 2014; Maturilli, 2020) that are part of the Baseline Surface Radiation Network (BSRN) (Driemel et al., 2018).

2.2 Reanalysis

We use data from the ERA5 reanalysis ~~(Hersbach et al., 2020)~~ (Hersbach et al., 2020; Graham et al., 2019) to place our findings into a large-scale context. The ERA5 reanalysis uses a fixed version of the ECMWF weather prediction model and data assimilation scheme to derive an observationally-constrained estimate of the atmospheric state. In-situ observations to constrain the model are scarce over the Arctic ocean. Satellite observations are more abundant, but often rejected by the assimilation scheme (Lawrence et al., 2019). As the underlying observational data sources change over time, reanalysis do not constitute a homogeneous record for estimating long-term trends. For example, the radiosonde measurements in Ny-Alesund have been assimilated into the reanalysis, but no corresponding measurements exist before 1993. In this paper, we primarily use ERA5 to confirm that changes seen in the local observations on interannual to decadal time scales are representative of developments over the central Arctic ocean.

2.3 CMIP6 models

We use data for the years 1993-2014 from the first ensemble member of the historical run of the models shown in table 1 for the CMIP6 DECK experiments (Eyring et al., 2016). Most models are in the mid-range or lower range of CMIP6 model resolution, but MPI-ESM-HR is a relatively high resolution model. All analyzed models allow for supersaturation with respect to ice in the sense that saturation with respect to ice is not a hardcoded humidity limit in the models' cloud schemes. In practice, the degree of possible supersaturation will strongly depend on the speed of removal of water vapour through depositional growth

of ice crystals. All models except GISS-E2-1-G have separate prognostic variables for liquid and frozen condensate. There is a mix of 1-moment microphysics schemes with fixed or diagnostic number concentrations of cloud particles, and 2-moment schemes with prognostic particle number concentrations.

To stratify the data based on large-scale flow conditions, we use 6-hourly atmospheric profiles on pressure levels, which substantially limits our model selection. Unless described otherwise, we use data from the model grid-point that is closest to Ny Ålesund.

Table 1. Evaluated climate models and their representation of mixed-phase microphysics. No. of phases refers to the number of condensate phases represented by prognostic variables. MIROC6 uses a 2-moment scheme for warm microphysics, but diagnostic ice number concentration. GISS uses a virtual mixed-phase scheme (vmp) to represent the effect of mixed-phase clouds within one prognostic variable for cloud condensate

model	atmosphere model	resolution	model documentation	no. of moments	no. of phases
AWI-ESM-1-1-LR	ECHAM6	T63 192 x 96 lon/lat	Stevens et al. (2013)	1-moment	2
GISS-E2-1-G	GISS-E2.1	2.5x2 degree; 144 x 90 lon/lat	Kelley et al. (2020)	'vmp'	1
MIROC6	MIROC6 AGCM	T85; 256 x 128 lon/lat	Tatebe et al. (2019)	1/2-moment *	2
MPI-ESM-1-2-HAM	ECHAM6-HAM	T63; 192 x 96 lon/lat	Neubauer et al. (2019a)	2-moment	2
MPI-ESM1-2-HR	ECHAM6	T127; 384 x 192 lon/lat	Stevens et al. (2013)	1-moment	2
MRI-ESM2-0	MRI-AGCM3.5	TL95; 192 x 96 lon/lat	Yukimoto et al. (2019a)	2-moment	2
HadGEM3-GC31-LL	MetUM-HadGEM3-GA7.1	N96; 192 x 144 lon/lat	Walters et al. (2019)	1-moment	2
ICON	ICON-AES	R2B9; 20 971 520 points, 5 km	Hohenegger et al. (2023)	1 and 2-moment	2

In climate models that use the same number of gridpoints on each latitude circle, the zonal grid spacing in the Arctic is much smaller than at lower latitudes. Nevertheless, the grid spacing in all models analysed here is too coarse to resolve the major topography of Svalbard (Figure 1), which reaches up to 1500 m. MPI-ESM-HR and GISS-E2-1-G have topographies that are at least twice as high as those of other models. In all models, but especially those with less resolved topography, the mountains stretch too far in the zonal direction, which is probably a result of smoothing the topographies for spectral models. While sub-grid parameterizations attempt to represent the effect of unresolved topography on the large-scale flow, the handover between resolved and unresolved processes is far from perfect, and at best reproduces the effects on the momentum budget (Sandu et al., 2019). Other orography-related effects such as precipitation or Foehn warming through downslope winds (Shestakova et al., 2022) can ~~hardly be well~~ at best be crudely represented in the CMIP6 models given their coarse representation of orography. In this paper, we focus on the representation of the ~~thermodynamics~~ thermodynamic properties of air masses exchanged between the Arctic and lower latitudes at larger scales rather than the local topographic effects, but the possibility that observations are affected by orography that is not adequately represented in the models must be kept in mind.

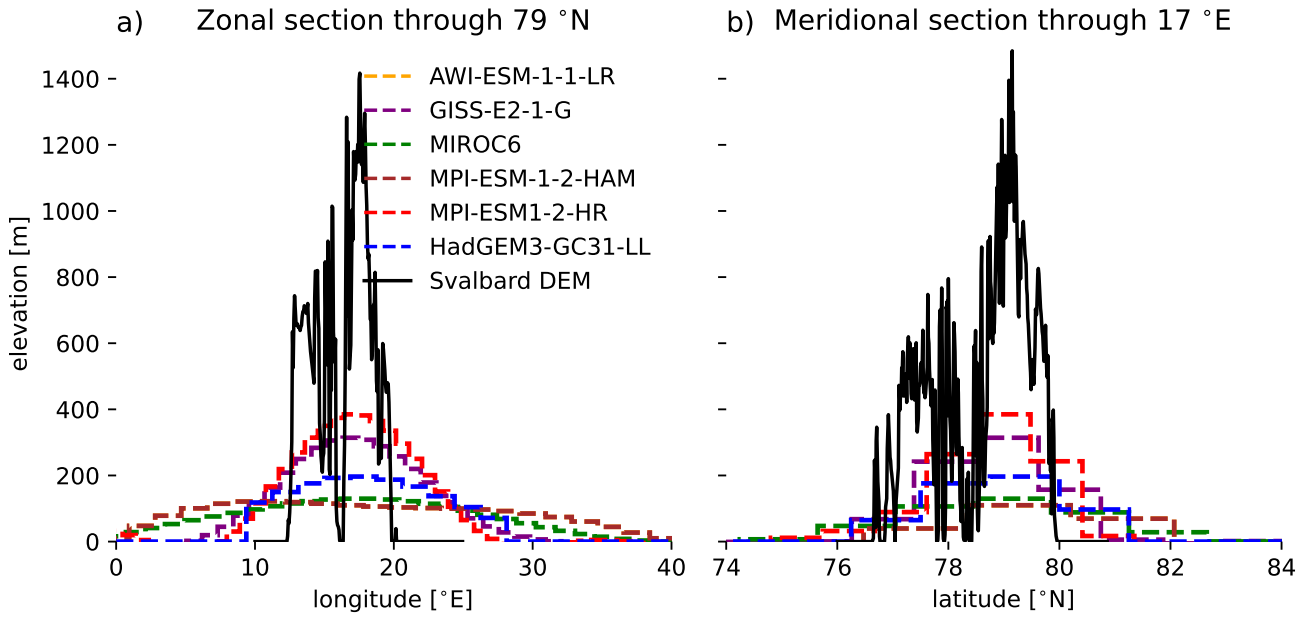

Figure 1. Zonal and meridional cross sections through the topography of Svalbard. AWI-ESM-1-1-LR and MPI-ESM-HAM use identical grids and topographies, hence the two lines exactly overlay each other on this plot. Svalbard DEM refers to the digital elevation model by Norwegian Polar Institute (2014)

2.4 global-storm resolving (km-scale) simulations with ICON

For some sensitivity analyses in section 2b, we also consider two simulations with the storm-resolving model ICON-Sapphire (Hohenegger et al., 2023) that apply a global quasi-uniform horizontal grid spacing of 5 km. The two simulations differ only by the applied representation of microphysical processes: in one simulation a one-moment scheme is predicting the specific mass of five hydrometeor categories (cloud water, rain, cloud ice, snow, graupel; Baldauf et al., 2011) and in the other simulation a two-moment scheme is predicting both the specific number and specific mass of six hydrometeor categories (cloud water, rain, cloud ice, snow, graupel, hail; Seifert and Beheng, 2006). Both schemes allow for excess water vapor with respect to ice saturation, i.e., supersaturation with respect to ice. The single-moment scheme calculates a diagnostic number of ice particles which depends only on temperature. It does not allow for homogeneous nucleation of ice particles from water vapor without any nuclei, but takes into account tendencies of cloud ice specific mass due to homogeneous freezing of cloud droplets and a simple heterogeneous nucleation rate. The two-moment scheme has a prognostic treatment of the number of ice particles and takes into account heterogeneous nucleation of ice particles (after Hande et al., 2015, including both immersion freezing and deposition nucleation), homogeneous freezing of cloud particles (after Jeffery and Austin, 1997), homogeneous nucleation of ice particles (after Kärcher et al., 2006) and Hallett-Mossop ice multiplication (after Beheng, 1982).

In addition to the parameterization of microphysics, radiation and turbulence are also parametrized but no shallow- or deep-convective parameterization is applied. The simulation setup closely follows the DYAMOND protocol (DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains; Stevens et al., 2019a). The model time step is 40 s and the vertical grid consists of 110 hybrid sigma height levels with a grid spacing of 400 m in the free troposphere gradually decreasing ~~toward~~ towards the surface and increasing toward the model top. Due to their high computational cost a simulation period of 10 days is covered by both simulations starting at 1 Feb 2020.

Analysis shown in this study is restricted to the last 5 days to allow for 5 days of spinup. This is a short period for ~~those used to~~ studying climatological effects, but it is long compared to the many case studies used to study the impact of microphysics on regional storm-resolving, or finer scale, simulations. The simulations are described in detail by Naumann et al. (2024) who show that for aggregated global statistics differences between the two simulations are typically larger than the day-to-day variability and hence the short simulation period is sufficient to identify systematic effects of changing the representation of microphysical processes. For our analysis, the difference between the ICON 1-moment and 2-moment schemes that occurs in the 5-day run is robust even for one-day averages ~~. Subsampling the CMIP models to five-day periods shows no evidence for similar shifts to arise from internal variability~~ (Figure S1).

2.5 Separating southwesterly and northerly flows

Throughout this paper, we stratify observations and model output based on the large-scale flow. To this end, we average the observed wind speeds between 650 and 775 hPa, i.e. well above the orography that constrains the local flow near Ny Ålesund (Maturilli and Kayser, 2017a; Schön et al., 2022). In CMIP6 models, we use wind speeds at the 700 hPa level. When this wind speed has both a westerly and southerly component, we classify the airmass as originating from the southwest, i.e. from the open ocean. When the wind speed has a northerly component, we classify the air as coming from the north, i.e. the largely sea-ice covered Arctic ocean, regardless of the zonal component of the flow. Differences between northeasterly and northwesterly advection (not shown) are small compared to those between northerly and southerly flows. We here do not show results for air masses coming from the southeast, i.e. the Barents sea, which has a more variable sea-ice cover.

3 Results

3.1 Temperature and moisture profiles

Air masses originating from the southwest are typically warmer than air masses from the north - the median temperature at 800 hPa is around -15 °C for southwesterly, but around -22 °C for northerly flows (Figure 2). The coldest and warmest (10th and 90th) percentiles indicate that cold air masses may arrive locally from the southwest and warm air masses from the north, as cold or warm air may be quickly circulated back, for example by smaller cyclones.

In contrast to other Arctic wintertime observations (Serreze et al., 1992) and most model results, no climatological temperature inversion emerges in the observed temperature profiles, neither in the median nor the percentiles. The Fjord has often

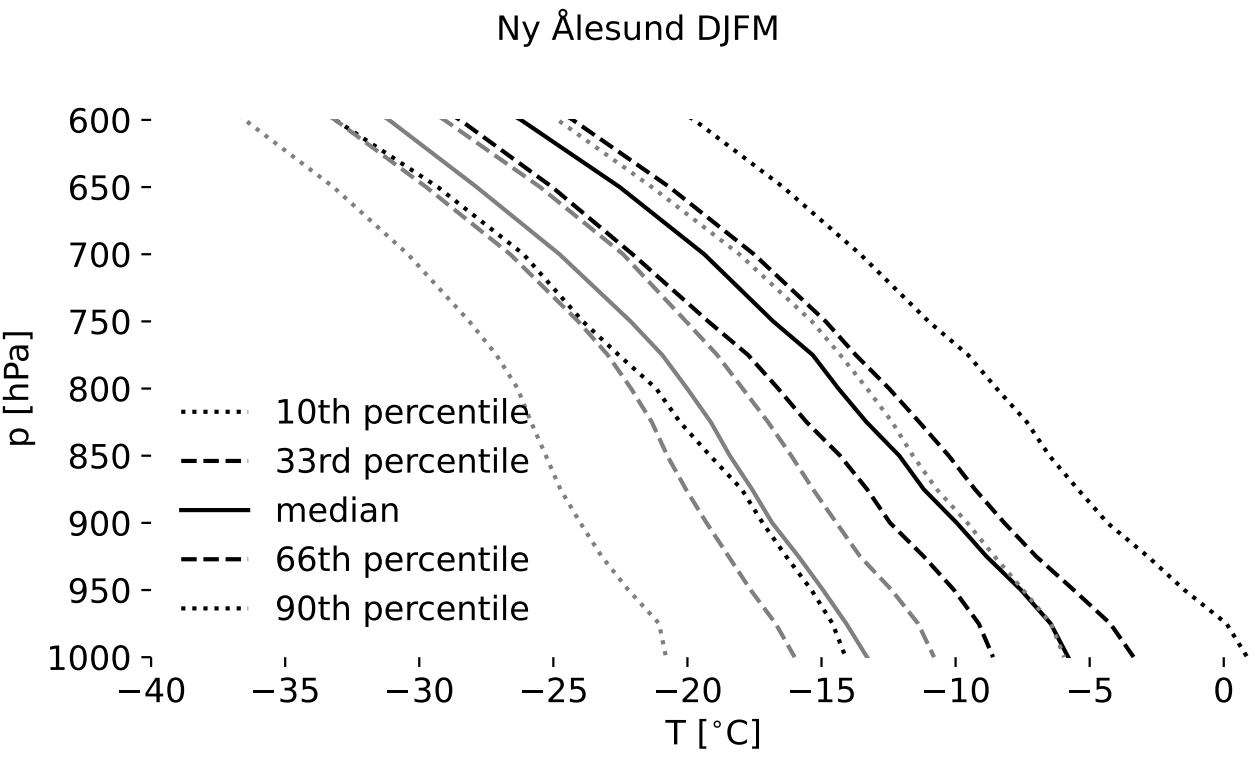

Figure 2. Temperature of air masses originating from the southwest (black) and north (gray lines) from radiosonde observations over Ny Ålesund for DJFM 1993-2014. Profiles of the 10th, 33rd, 50th (median), 66th, and 90th percentile of temperature at each level are shown.

remained ice-free throughout the winter in the years covered by our dataset, and radiosonde observations from ~~Danmarkshaven~~ Danmarkshavn in Greenland (~~not shown~~analysis not shown, data available through IGRA, Durre et al. (2018)) suggest that the presence of upwind topography can cause additional mixing and destroy a temperature inversion that may have existed in the arriving air mass. We ~~refrain from interpreting~~ therefore do not interpret the deviation of modelled and observed temperatures at lower levels (marked by gray shading in profile plots) as a model bias~~, as the~~. These observations might simply not be representative of the scales the models attempt to describe.

As most CMIP6 models do not provide the high-frequency data required to compute these flow-based profiles, we initially evaluate the monthly mean data (Figure 3). In the free troposphere, a few models are warmer than observations by 2-3 degrees, with one model having a warm bias around 5 degrees. Most models have a cold bias of up to 5 degrees, and some are cold-biased by nearly 10 degrees. The models for which high-frequency data is available (coloured lines) mostly have modest cold biases of 2-3 degrees, but MPI-ESM-HAM closely matches observations and GISS-E2-1-G is among the models with the strongest cold biases around 10 degrees. The tendency for a cold bias in the subset of models is thus representative for the CMIP6 ensemble, but models with a warm bias (which are few) and those with stronger, but not extreme cold biases around 5 degrees are not

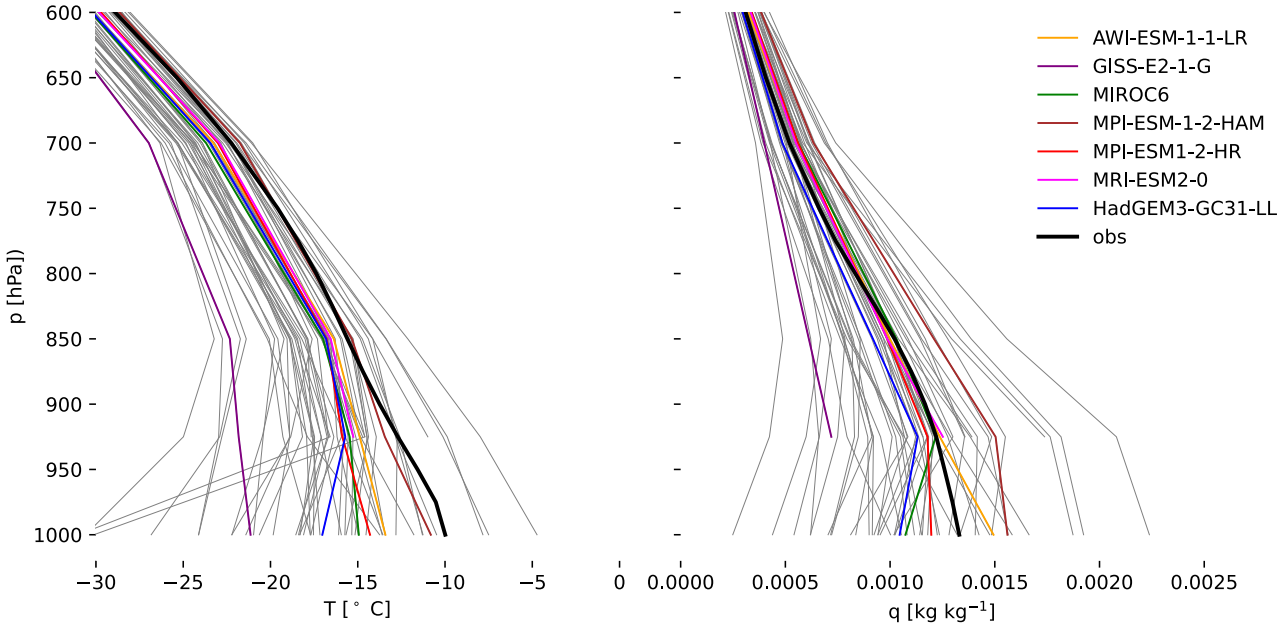

Figure 3. a) Temperature and b) specific humidity profiles from radiosonde observations over Ny Ålesund and monthly mean output from CMIP6 models for DJFM 1993-2014. Coloured models are those that will be analyzed in the remainder of the manuscript using high-frequency data.

represented in this subset. For the mean specific humidity profiles, most models are close to observations, but GISS-E2-1-G (dry) and MPI-ESM-HAM (moist) span a substantial part of the CMIP6 ensemble spread. In the following, we focus on analysing northward and southward moving air masses in the subset of models with high-frequency data, bearing in mind that they represent the overall cold bias, but not every aspect of the full CMIP6 ensemble.

In the free troposphere, i.e. above 850 hPa, models tend to be colder than observed (Figure 4). This cold bias is somewhat more pronounced in northerly than in south-westerly flows, and generally stronger for the coldest quantiles. ~~For models with a cold bias in global mean temperature, this might reflect an Arctic amplification of model biases in parallel to the Arctic amplification of climate change.~~ The tendency towards a cold bias over Svalbard is consistent with the known near-surface cold bias of CMIP6 models over the central Arctic ocean in winter (Davy and Outten, 2020).

AWI-ESM-1-1-LR matches observed temperatures for the warmest (90th) quantile in air masses flowing poleward in the lower troposphere, and only develops a cold bias on the order of 1 K at 600 hPa. The warmest air masses coming from the south do not show any signs of surface decoupling or ~~the formation of~~ temperature inversions in either observations (Figure 2) or the model (not shown), such that the specific local conditions of the Fjord that can cause additional mixing do not lead to a mismatch between observed and ~~modelled~~ AWI-ESM-1-1-LR profiles under these conditions. For colder air masses, especially those advected from the north, AWI-ESM-1-1-LR has a cold bias on the order of 1-2 K.

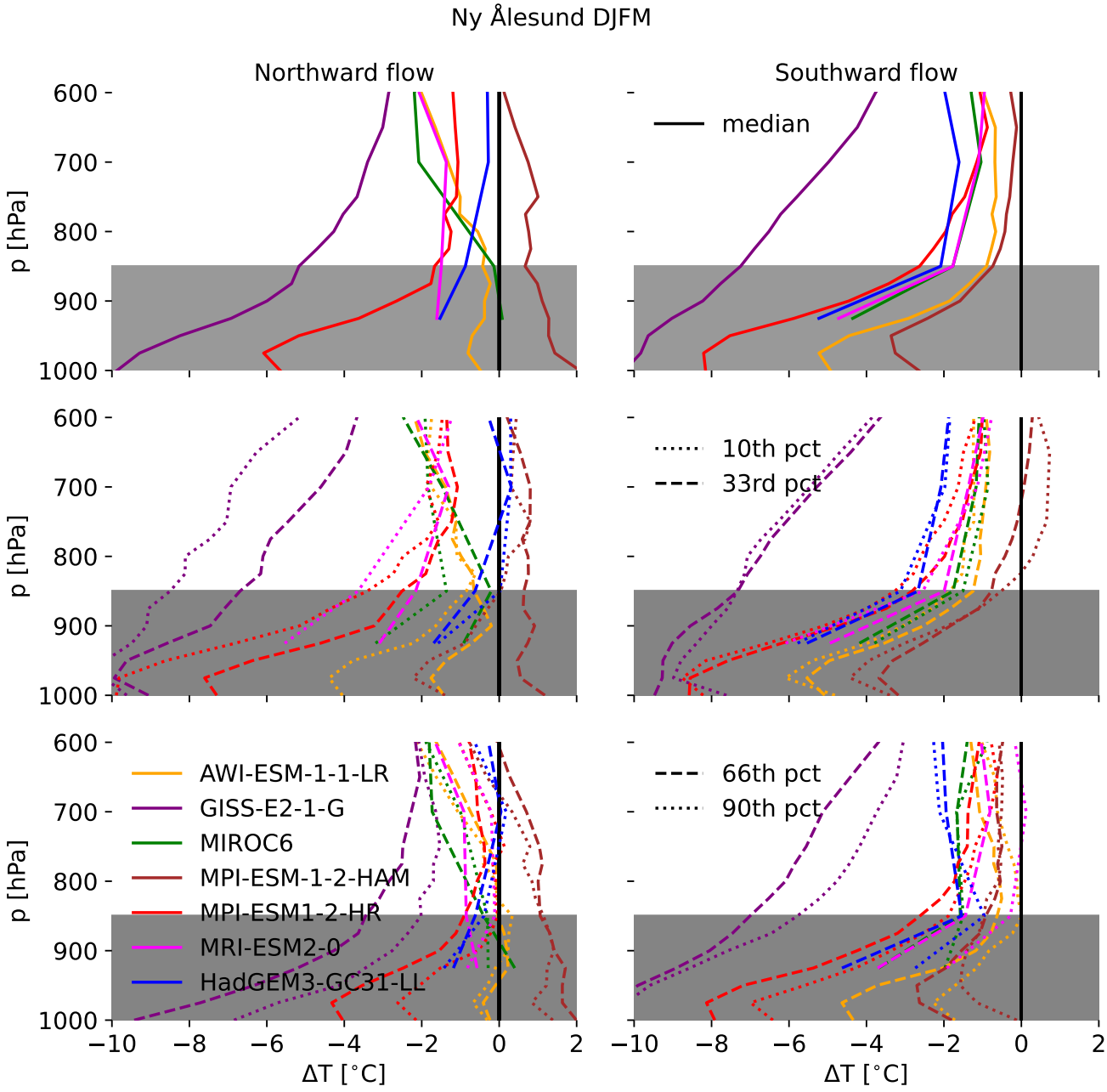

Figure 4. Temperature biases of models against radiosonde observations for air masses originating from the southwest (left) and north (right) over Ny Ålesund for DJFM (1993-2014). Biases are shown for the 50th (median, first row), 10th, 33rd, (second row), 66th, and 90th (third row) percentile of temperature at each level. Gray shading marks the altitudes at which we expect strong effects of local topography on the observations that are not represented in the models.

GISS-E2-1-G has the most pronounced cold bias among the models analysed here, ranging from 1-2 K for the warmest air masses in poleward flows to 5-8 K for air masses coming from the North. The global mean temperature in GISS-E2-1G is largely unbiased with a compensation between a warm bias at low and southern high latitudes and a pronounced Arctic cold bias (Kelley et al., 2020), suggesting that the cold bias seen in cold Arctic air masses is not just an amplification of a global bias, but originates in mid- to high latitudes of the Northern hemisphere.

MIROC6 has a cold bias on the order of 1-2 K, which is mostly constant across poleward and equatorward flows and temperature quantiles. Globally, the model is biased warm even when comparing the preindustrial simulation to more recent observations (Tatebe et al., 2019).

MPI-ESM-HAM matches observed temperatures remarkably well in both southwesterly and northerly flows and for all temperature quantiles. In contrast to some of the other models, MPI-ESM-HAM only produces marked temperature inversions in air masses coming from the north.

MPI-ESM-1-2-HR is virtually unbiased for the warmest air masses advected from the south, but has a cold bias of at least 2 K above the temperature inversion for the coldest air masses.

MRI-ESM2-0 matches observed temperatures for the warmest air masses advected from the south and has a cold bias of 1-2 K above the modelled temperature inversion for the coldest air masses.

HadGEM3-GC31-LL has a weak cold bias in air masses advected from the south, and a cold bias of at least 2 K for air masses advected from the North, with a stronger cold bias for colder quantiles.

Specific humidity in most percentiles increases rapidly towards the surface between 700 and 850 hPa, whereas the increase towards the surface is weaker below 900 hPa (Figure 5). Near the surface, the moistest poleward-moving air masses contain more than 0.003 kg kg $^{-1}$ of moisture, roughly twice as much as the moistest airmassses drifting southward. The difference in median air masses exceeds a factor of two.

Biases in specific humidity (Figure 6) partly parallel those in temperature, but as for the monthly mean profiles (Figure 3), the models do not display a consistent dry bias as could have been expected based on the temperature biases discussed above. Again, the mismatch between observed and modelled profiles in the boundary layer might be due to local conditions in the Fjord that do not reflect the large-scale average represented in a model grid box.

AWI-ESM-1-1-LR has a slight dry bias in warm air masses, but tends towards a moist bias in cold air masses. GISS-E2-1-G has less humidity than observed under all conditions, consistent with the model's pervasive cold bias discussed above. MIROC6 matches the moisture content of warmer air masses rather well, but has a moist bias in cold air masses which is more pronounced than that in AWI-ESM-1-1-LR. Note that MIROC6 has a cold bias under these conditions, suggesting that these air masses must be substantially more saturated in the model than observed in Ny Ålesund (see the high share of saturated air masses at cold temperature in MIROC6 in Figure 8). MPI-ESM-HAM has a tendency towards a moist bias under all conditions, which is more pronounced for the coldest quantiles of air masses being advected equatorwards. MPI-ESM1-2-HR matches observed humidities in the free troposphere rather well. As discussed above, an apparent dry bias in the boundary layer might be due to non-representative conditions in the Fjord.

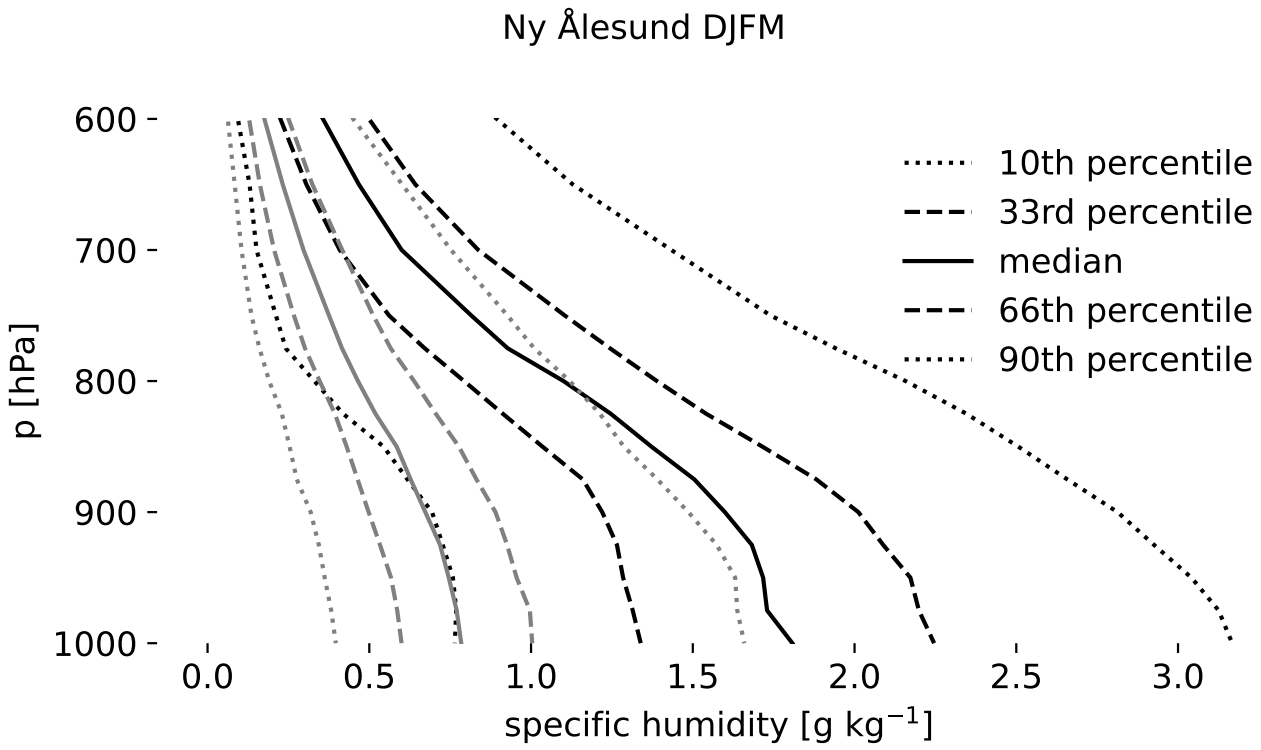

Figure 5. Specific humidity of air masses originating from the southwest (black) and north (gray lines) from radiosonde observations over Ny Ålesund for DJFM 1993-2014. Profiles are shown for the 10th, 33rd, 50th (median), 66th, and 90th percentile of specific humidity at each level.

3.2 Relative humidity

To evaluate the relationship between temperature and specific humidity - effectively the relative humidity - in models, we plot specific humidity against temperature (Figure 7). This avoids any ambiguity about using relative humidity with respect to water or ice, and has the added advantage of displaying temperature and humidity at the same time. Below 850 hPa, where most measurements will be influenced by the local boundary layer (Maturilli and Kayser, 2017a), specific humidity is constrained by saturation specific humidity with respect to ice, with substantial variability below that value. This variability is in contrast to measurements from the MOSAiC campaign, where the wintertime boundary layer is much closer to saturation with respect to ice (not shown). We therefore restrict our model evaluation to the lower free tropospheric levels between 600 and 700 hPa, where observed humidity tends to be either close to or above saturation with respect to ice, and bounded by saturation with respect to water, or substantially below saturation~~(Figure 8)~~.

Most models are close to saturation with respect to ice for cold temperatures, and somewhat below saturation with respect to ice at warmer temperatures (Figure 8). In particular at cold temperatures, models tend to lack ~~both supersaturated and~~

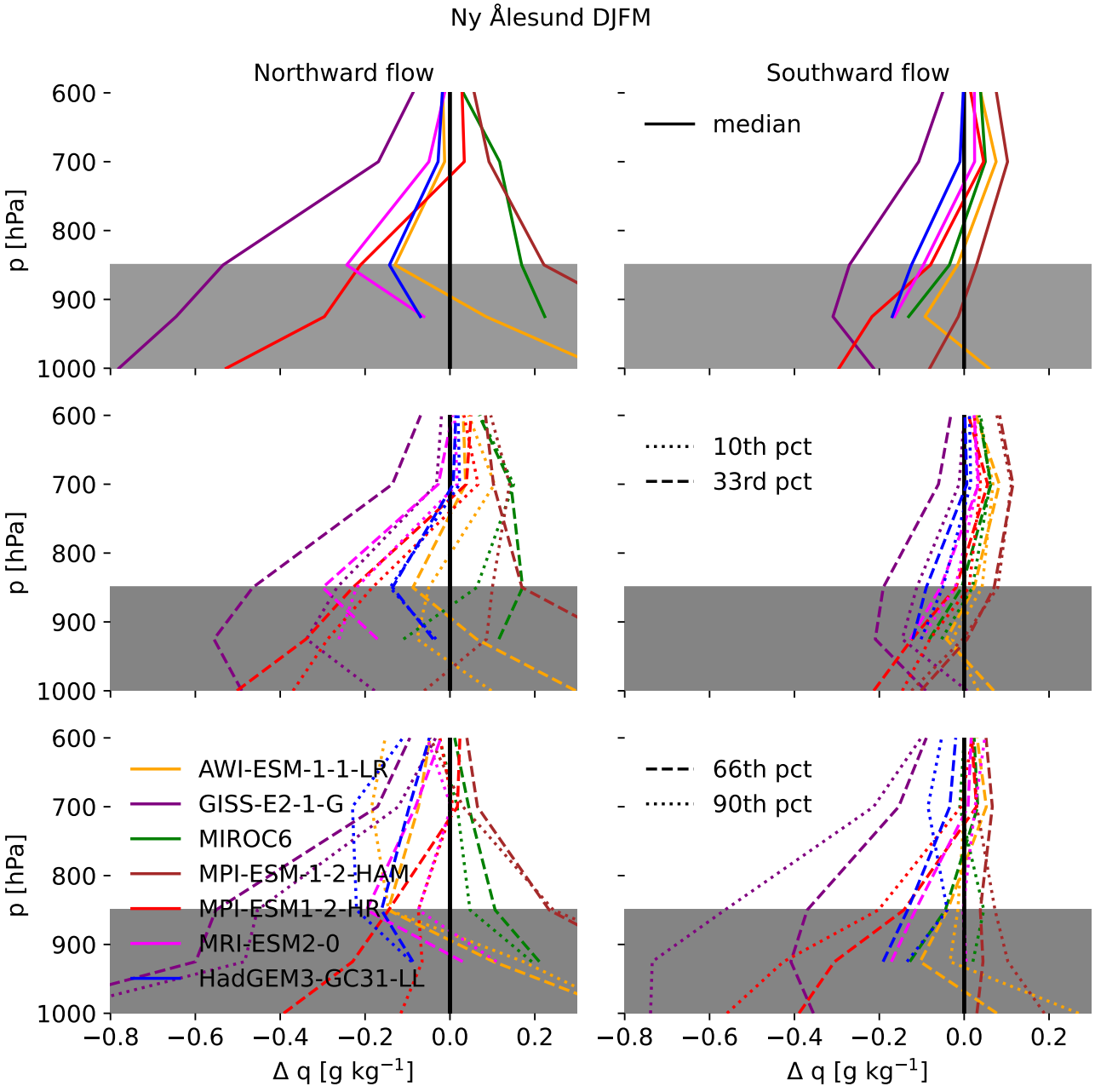

Figure 6. Specific humidity biases of models against radiosonde measurements for air masses originating from the South-West (left) and North (right) over Ny Ålesund for DJFM. Biases are shown for the 50th (median, first row), 10th, 33rd, (second row), 66th, and 90th (third row) percentile of specific humidity at each level. Gray shading marks the altitudes at which we expect strong effects of local topography on the observations that are not represented in the models.

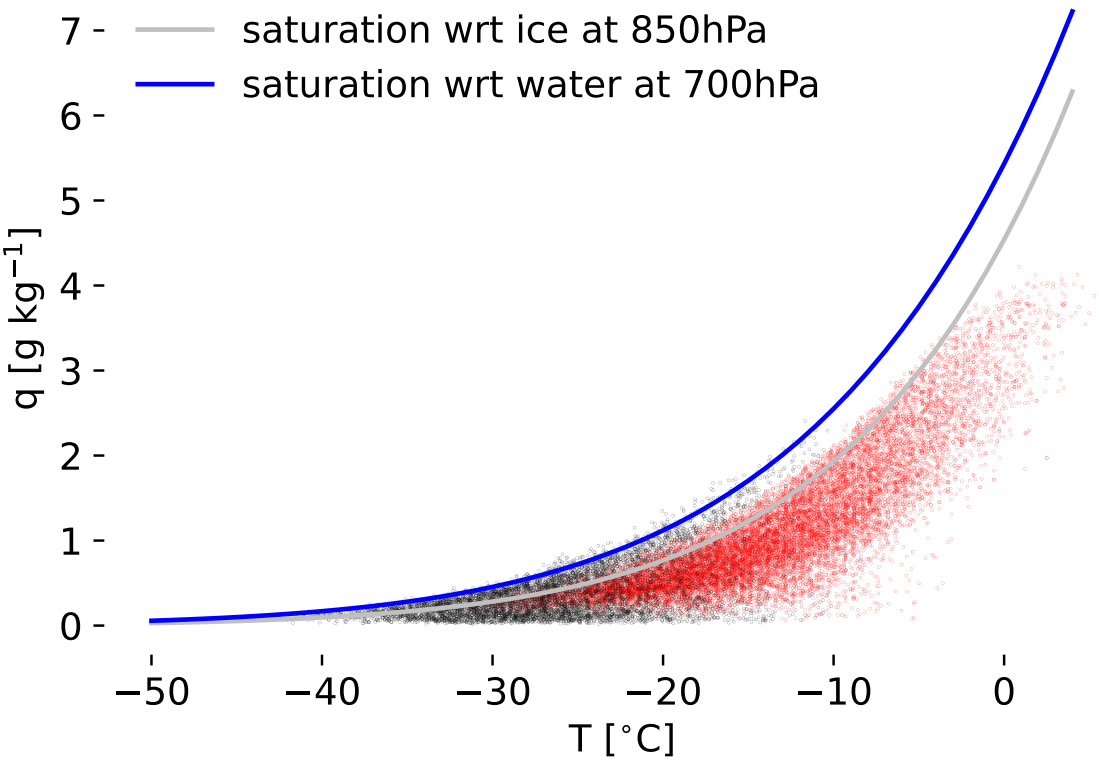

Figure 7. Specific humidity vs. temperature in radiosonde measurements over Ny Ålesund for DJFM. Each dot represents a single measurement at one time and interpolated to a CMIP pressure level.

~~substantially subsaturated~~ supersaturated conditions frequently seen in observations~~, both for poleward and equatorward flows~~. One exception that we will discuss further below is the MRI-ESM2-0 model which more frequently reaches humidities close to saturation with respect to water, and thus supersaturation with respect to ice, at cold temperatures below -15 °C. Several models also lack strong subsaturation with respect to ice that is seen in observations.

Larger spread in observed than modelled specific humidities for a given temperature is to be expected - the radiosonde measurements reflect local conditions right at the sensor of the sonde, whereas the model value is supposed to represent a grid-box average. Sub-gridscale fluctuations of humidity are substantially larger than those of temperature, and accounted for in cloud schemes of large-scale models that produce clouds even when the grid-box average is well below saturation (Sundqvist, 1978). We use km-scale model runs following the DYAMOND protocol (Stevens et al., 2019b) to investigate to what extent models better capture the observed distribution of relative humidity with respect to ice at higher resolution, and vary the microphysics scheme to examine the role of parameterized physics for the distribution (Figure 9).

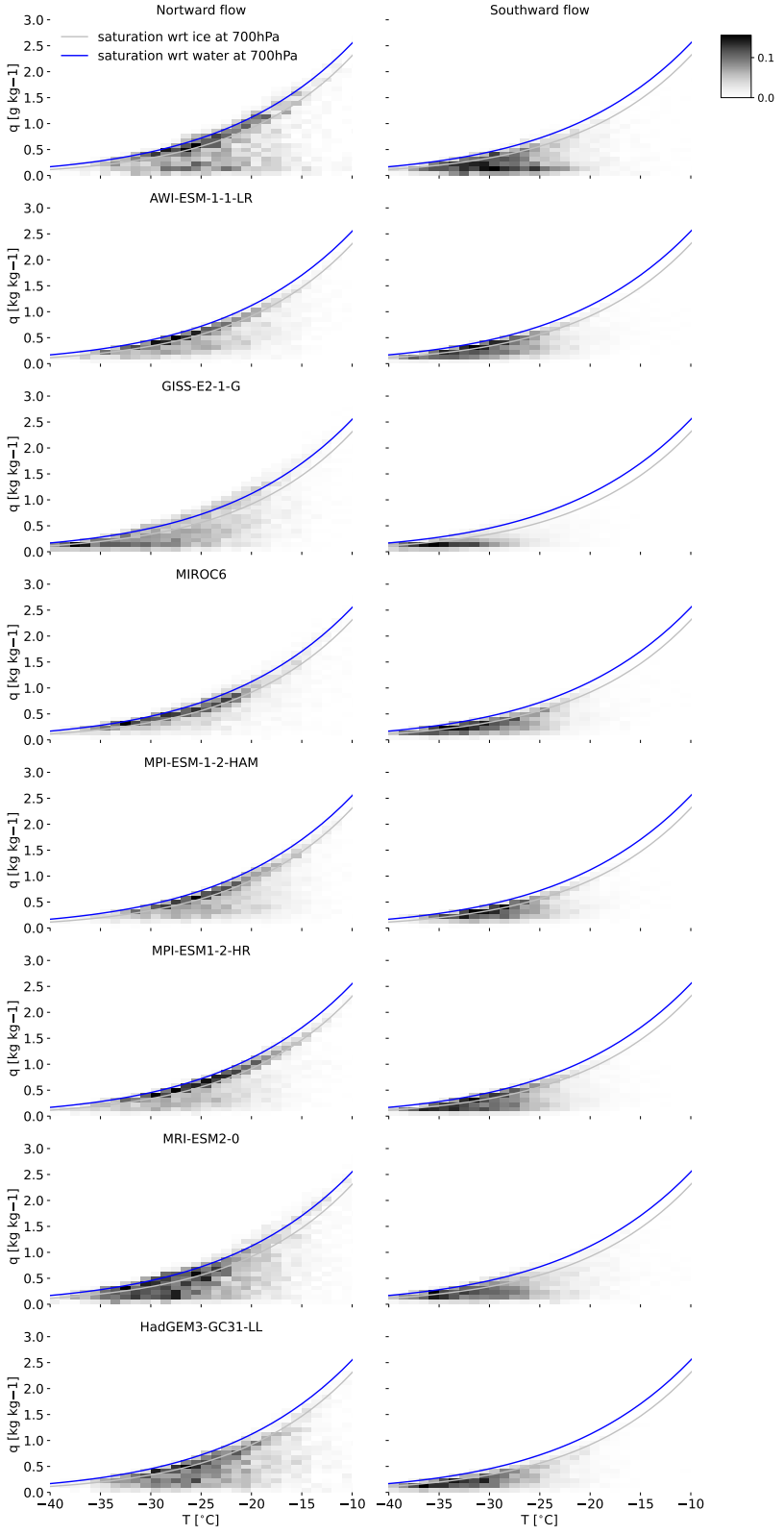

Figure 8. ~~Specific~~ Bivariate pdf of specific humidity vs. temperature in air masses originating from the South-West (left) and North (right) over Ny Ålesund for DJFM in radiosonde observations (top row) and climate models.

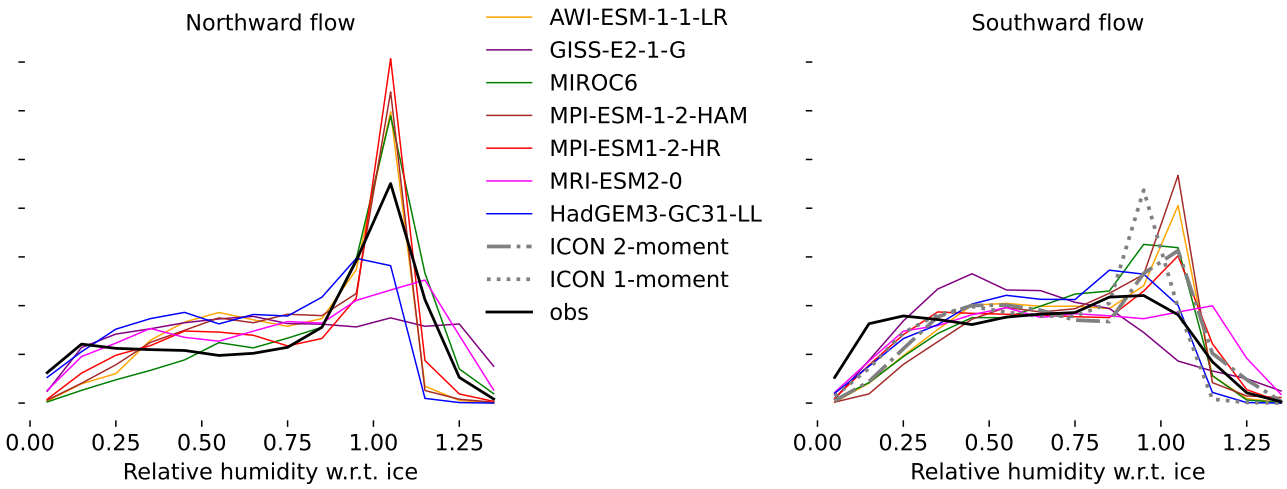

Figure 9. PDF of saturation with respect to ice for CMIP models and observations over Ny Ålesund in air masses originating from the South-West (left) and North (right) during DJFM. ICON data (right) are from a much shorter run than CMIP models and cover the central Arctic ocean (70° N to 90° N) to obtain useful statistics. They are plotted alongside observations and CMIP model output for air masses originating from the North as they represent air mass properties in that source region.

In both southwesterly and northerly flows, observed relative humidity with respect to ice is most frequently below 30 or close to 100 %~~, with some indication of a weaker intermediate mode~~. Near-saturated conditions are more frequent in southwesterly than northerly flows. Most CMIP models underestimate the occurrence of strong undersaturation and of supersaturation with respect to ice, consistent with the above results for specific humidity against temperature.

The lack of supersaturation with respect to ice in the km-scale run is qualitatively similar to the CMIP models when using 1-moment microphysics, but improves when switching to the 2-moment scheme. The CMIP ~~models~~ model that most frequently simulates supersaturated conditions under both southwesterly and northerly flows, MRI-ESM2-0, is also one of the few that employs a 2-moment microphysics scheme. MPI-ESM-HAM does not substantially outperform MPI-ESM-HR or AWI-ESM-LR in this respect, despite having largely the same physics with the exception of a 2-moment microphysics scheme. The GISS model overestimates supersaturation in southwesterly flows and subsaturation in northerly flows.

These results suggest that 2-moment microphysics schemes may have an advantage in correctly representing the observed distribution of humidity at and above ice saturation in the Arctic free troposphere. We hypothesize that 2-moment schemes can produce low ice number concentrations despite being supersaturated with respect to ice, slowing down the removal of water vapour by depositional growth of ice crystals. In contrast, 1-moment schemes would generally tend to diagnose high ice number concentrations under ice supersaturation, which will lead to quick removal of water vapour.

Ice supersaturation is indeed associated with substantially lower ice crystal plus snow number concentrations in the ICON 2-moment scheme than in the 1-moment scheme (Figure 10, note the logarithmic y axis). We compare the sum of snow and ice

number concentrations, as both hydrometeor species contribute to the depletion of supersaturation and the partitioning between both is not consistent between the different microphysics schemes.

Two-moment microphysics schemes thus have a structural advantage over one-moment schemes in representing supersaturation with respect two ice. Other factors in models can also influence the existence or lack of ice supersaturation. For example, the assumption that in a cold, cloudy grid box, water vapour cannot exceed its saturation value, is hardcoded in some models (Tompkins et al., 2007). Such models might still show supersaturated values in the above pdf due to vertical averaging.

The km-scale ICON runs produce a dry mode with relative humidity with respect to ice below 50 % that is more similar to observations than the much coarser CMIP models. Comparing the somewhat higher resolution model MPI-ESM-HR to the physically similar but coarser models MPI-ESM-HAM and AWI-ESM-LR also suggests that higher horizontal resolution might be helpful to represent low relative humidities as observed. However, the ICON runs shown here only cover a few days, and in contrast to the improvement of ice supersaturation with the 2-moment scheme, a better representation of the dry mode is not backed up by mechanistic understanding, and is regionally less robust (not shown). Whether and why high-resolution models better capture the low relative humidity values should be investigated in future research.

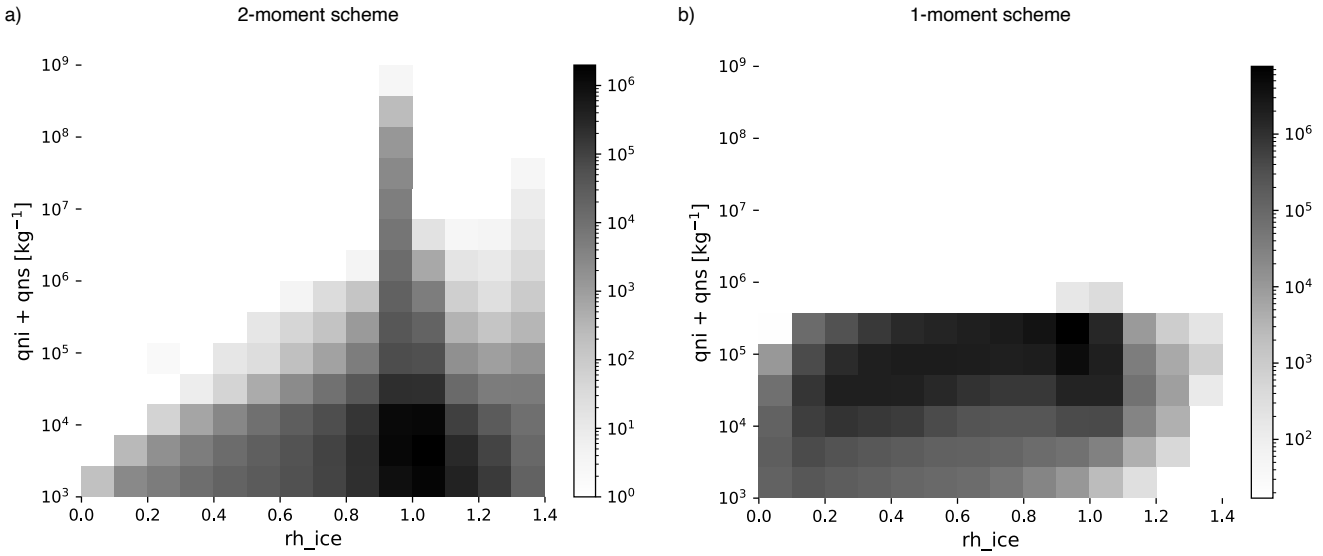

Figure 10. ~~Cloud~~ Bivariate pdf of cloud ice plus snow number concentration against saturation with respect to ice in the Arctic north of 70° N a) in the ICON 2-moment setup and b) diagnostic number concentrations in the 1-moment setup. Note the logarithmic y-scale and color scale.

3.3 Longwave radiation

The observed wintertime net longwave radiation in Ny Ålesund (Figure 11, defined positive downwards) shows the bimodal distribution typical of Arctic winter, with a cloudy mode characterized by the presence of cloud liquid water and net surface

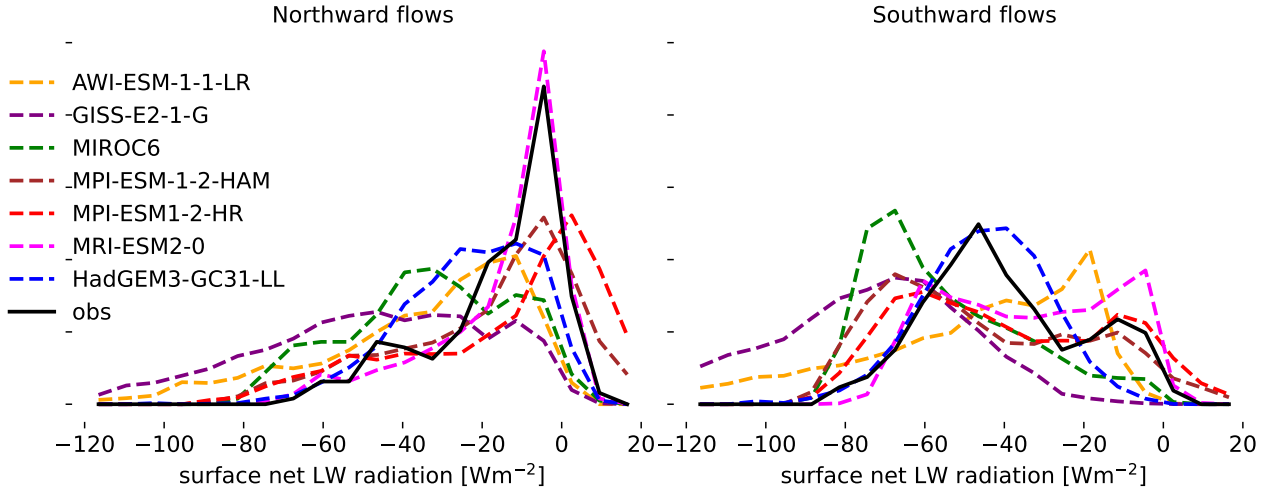

Figure 11. PDF of net longwave radiation (positive downwards, i.e. towards the surface) in air masses originating from the South-West (left) and North (right) over Ny Ålesund for DJFM.

radiation between 0 and -10 Wm^{-2} and a clear mode with longwave radiative cooling around -40 Wm^{-2} (Stramler et al., 2011). The cloudy mode is dominant in air masses arriving from the southwest, confirming that is is caused by the advection and transformation of warm, moist air masses, whereas the clear mode plays a more important role in air masses arriving from the central Arctic ocean. Consistent with analyses over the central Arctic ocean (Duffey et al., 2024), not all models appear to reflect this bimodality. But model-observation differences in net radiation may be difficult to interpret, as the observational sensor is located over land, whereas some models have a substantial ocean part in the nearest grid box.

We therefore focus on the downward longwave radiation for further evaluating the models (Figure 12). In air masses originating from the southwest, models with strong cold biases in the lower troposphere also tend to understimate downward longwave radiation. But in air masses coming from the north, some models strongly overestimate downward longwave radiation despite being biased cold in lower tropospheric temperatures, in particular the GISS model and AWI-ESM-1. As this bias does not seem related to the typical temperature profiles of air masses in these models, it has to be associated with the radiative properties of the atmosphere, most likely related to clouds.

Plotting temperature at the lowest atmospheric level (925 hPa) against the downward longwave radiation at the surface (Figure 13) shows two linear relationships corresponding to the bimodal distribution of surface net longwave radiation, with clear-sky emissions at the lower end and fully opaque clouds at the upper end of the distribution. The likelihood of measurements being close to the lower bound, i.e. representing radiatively clear boundary layers, increases at colder temperatures, with the clear state becoming dominant around 265 K. AWI-ESM1-1 appears to show values of downward longwave radiation that exceed the ~~upper bound of cloudy conditions for~~ near-linear relationship formed by clouds radiating as blackbodies at a given temperature. We attribute this to the model having its closest gridbox over open ocean, whereas all other models for which the

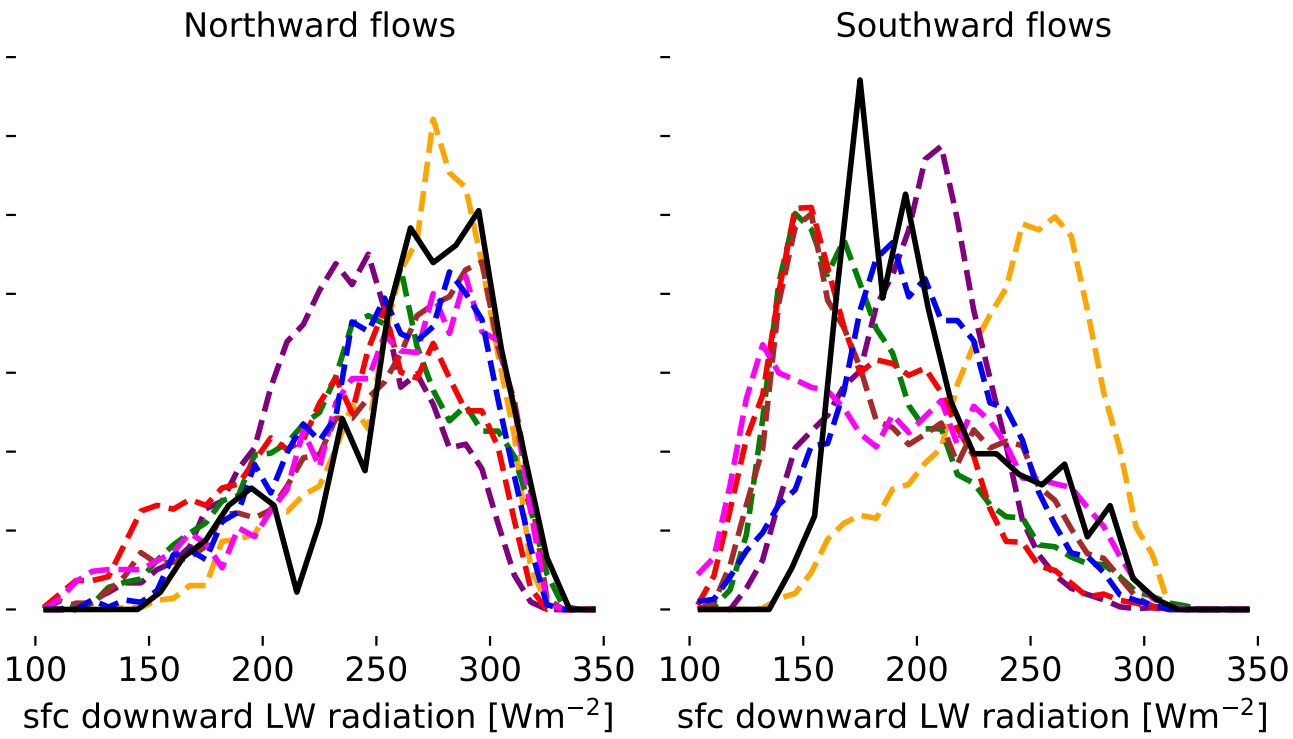

Figure 12. PDF of downward longwave radiation at the surface in air masses originating from the South-West (left) and North (right) over Ny Ålesund for DJFM. Line colors as in Figure 11

land sea-mask was available have a fully or mostly land-covered grid box close to Ny Ålesund~~(see land-sea fractions in Figure 2)~~. When choosing the next grid box to the ~~west~~ east in AWI-ESM1-1, which is land-covered, the points above the upper bound no longer appear (Figure 14).

In GISS-E2-1-G, and to a lesser extent in MPI-ESM-HR, the cloudy state of the boundary layer, i.e. values close to the upper bound of downward longwave radiation, are more frequent than observed at temperatures below 265 K. High-emissivity, usually liquid-containing clouds are thus more frequent in these models than in observations at cold temperatures. This is consistent with the finding of Kelley et al. (2020) that the virtual mixed-phase cloud scheme in GISS leads to an overestimation of supercooled liquid compared to satellite observations. For a given tropospheric temperatures, more emissive clouds lead to warmer surface temperatures in the Arctic, but they also lead to more efficient radiative cooling of tropospheric air, which could contribute to the strong tropospheric cold bias in GISS-E2-1-G.

While past climate model evaluations often found a lack of supercooled liquid water and high emissivity clouds at high latitudes (Cesana et al., 2012; Komurcu et al., 2014), we here see that some CMIP6 models maintain higher cloud emissivity under cold conditions than observed in Ny Ålesund. This is in agreement with Cesana et al. (2022), who showed that CMIP6

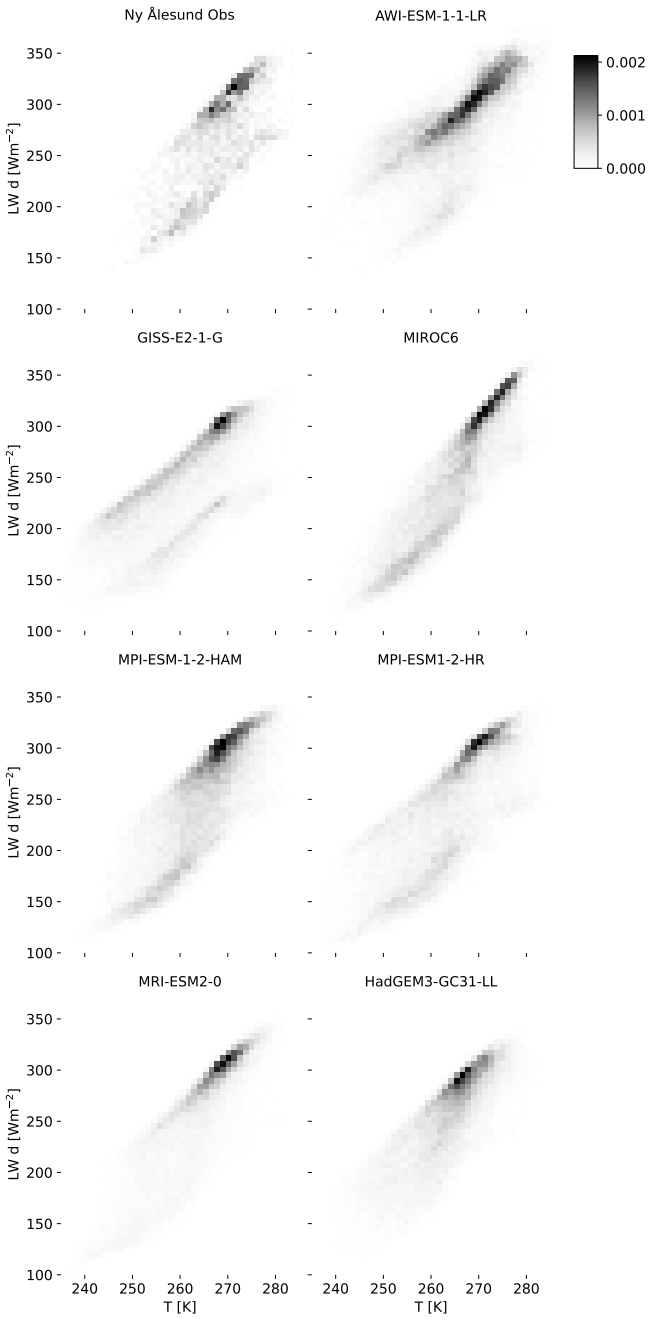

Figure 13. ~~Temperature~~ Bivariate pdf of temperature at 925 hPa vs downward longwave radiation at the surface over Ny Ålesund for DJFM.

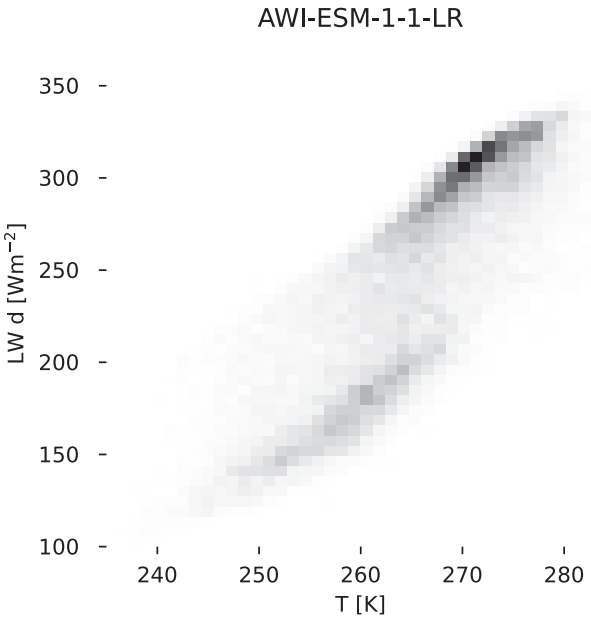

Figure 14. ~~Temperature~~ Bivariate pdf of temperature at 925 hPa vs downward longwave radiation at the surface over Ny Ålesund for DJFM using ~~an ocean~~ a land gridpoint next to NYA for AWI-ESM-1-1-LR.

models with a simple, temperature-dependent phase partitioning overestimate liquid in mixed-phase clouds over the Southern Ocean.

3.4 Precipitable water and representativity

Integrating precipitable water from the profiles measured and modelled in southwesterly and northerly flows (Figure 15) shows that GISS-E2-1G, which is generally cold-biased, has a mode at substantially lower values of precipitable water than observed in southwesterly flows. ~~MIROC6 appears to underestimate the frequency of occurrence of high values of precipitable water, but such a bias is not evident in the profiles of precipitable water and may well be an artefact of the model not reporting data at the 1000 hPa pressure level, which leads to a low bias when integrating over the atmospheric column.~~ AWI-ESM-1-1, MPI-ESM-HAM and MPI-ESM-LR model a realistic distribution of precipitable water in southwesterly flows (Figure 15 a).

In northerly flows (Figure 15 b), GISS-E2-1G ~~and MIROC6 produce modes that are~~ produces a mode that is lower than observed, and MPI-ESM-HR has a realistic mode but a lower frequency of occurrence for moister air masses. MPI-ESM-HAM and AWI-ESM1-1 have realistic distributions of precipitable water.

Annual mean values of precipitable water in southwesterly and northerly flow correlate very well between the gridpoint next to Ny-Alesund and a section across Fram Strait in models, both across models and for the interannual variability (Figure 16). The long-term observations of atmospheric profiles at AWIPEV are thus suitable to evaluate air-mass properties at this crucial gateway to the Arctic in climate models and to record important trends and year-to-year variations in moisture content.

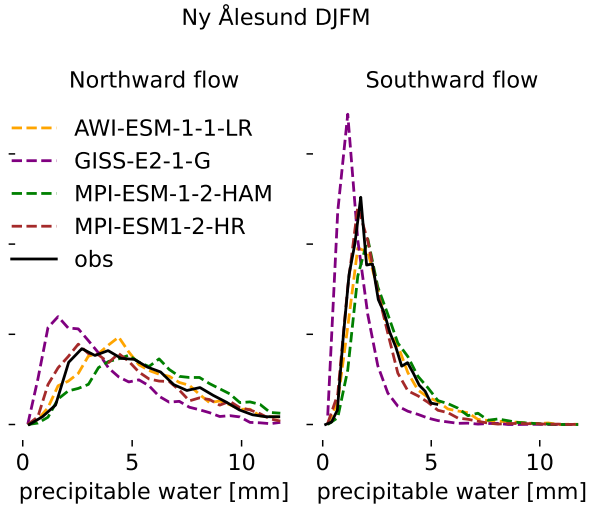

Figure 15. Histogram of modelled and observed values of precipitable water in southwesterly (left) and northerly (right) flows over Ny Ålesund for DJFM. Models that do not provide data at 1000 hPa are omitted.

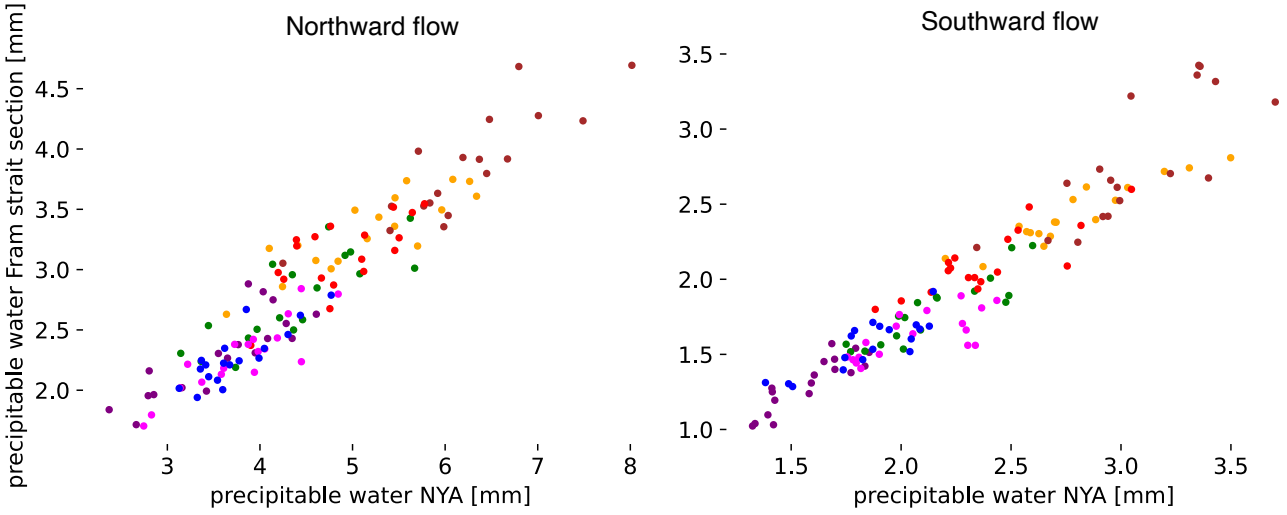

Figure 16. DJFM annual mean of precipitable water modelled for the grid point closest to Ny Ålesund vs. mean over a section across Fram Strait for southwesterly (left) and northerly flows (right). Different colors decode different models following the color coding in Figure 15.

3.5 ~~Trends~~Observed trends

The annual mean precipitable water in air masses advected over Ny Ålesund during DJFM from the South-West has a strong year-to-year variability and an increasing trend of 0.036 mm yr^{-1} with a p-value of 0.056. The 30-year trend (1993-2022) is thus statistically not significant to the standard threshold of 0.05 (Figure 17 a). In air masses advected from the North, a sudden shift towards moister conditions after 2004 stands out from the year-to-year variability. The linear trend is 0.027 mm yr^{-1} with a p-value of less than 0.002, but a linear regression is obviously a poor description of the observed step change. This rapid Arctic wintertime warming and moistening in the early 2000s (Dahlke and Maturilli, 2017) can also be seen in ERA5 precipitable water averaged over the polar cap north of 70 $^{\circ}N$ (Figure 17 b).

The year 2020 stands out as particularly dry for both southwesterly and northerly flows - it is among the two years with the lowest precipitable water values in southwesterly flows, and in line with low values of the 1990ies and earlier 2000s in northerly flows. This is in line with the overall meteorological conditions encountered during the MOSAiC expedition between Fall 2019 and Summer 2020, which were dominated by strong zonal flows and little meridional advection which could bring moist air masses to the central Arctic (Lawrence et al., 2020). Measurements in Ny Ålesund thus reflect the changing conditions over the central Arctic ocean, including the particular conditions during 2020.

The shift towards moister winter conditions over the central Arctic has been attributed to a more frequent occurence of southerly winds over Fram Strait (Dahlke and Maturilli, 2017; Nygård et al., 2020).

4 Summary and conclusions

We compare temperature and moisture in the lower Arctic troposphere between CMIP6 climate models and observations at Ny Ålesund, Svalbard, which is located in a major gateway for air-mass exchanges between the Arctic and lower latitudes. We focus on the lower free troposphere, as boundary-layer conditions at the measurement site are influenced by the local topography that is not resolved by large-scale models. Climate models tend to be cold-biased (Davy and Outten, 2020) especially for the coldest temperatures which could be caused by thermodynamic biases or biases in the atmospheric circulation that cause too little exchange of air masses and thus too long residence time of air within the Arctic. Air masses entering the Arctic tend to be less biased or even unbiased in their temperature structure than those leaving the Arctic, and Winkelbauer et al. (2024) show that CMIP6 models overestimate atmospheric energy convergence into the Arctic in winter, which suggests that thermodynamic biases are the more likely cause.

Relative humidity in models is close to saturation over ice more frequently in models than in observations. ~~Models lack both~~ Most models lack supersaturation with respect to ice and several models also lack subsaturation at relative humidities with respect to ice around or below 30 %. Models with 2-moment microphysics that compute the ice number concentration instead of assuming it better represent supersaturation and associate it with lower ice number concentrations. We see some indication that high-resolution models better represent strong subsaturation, but whether and why this is a robust effect remains to be investigated.

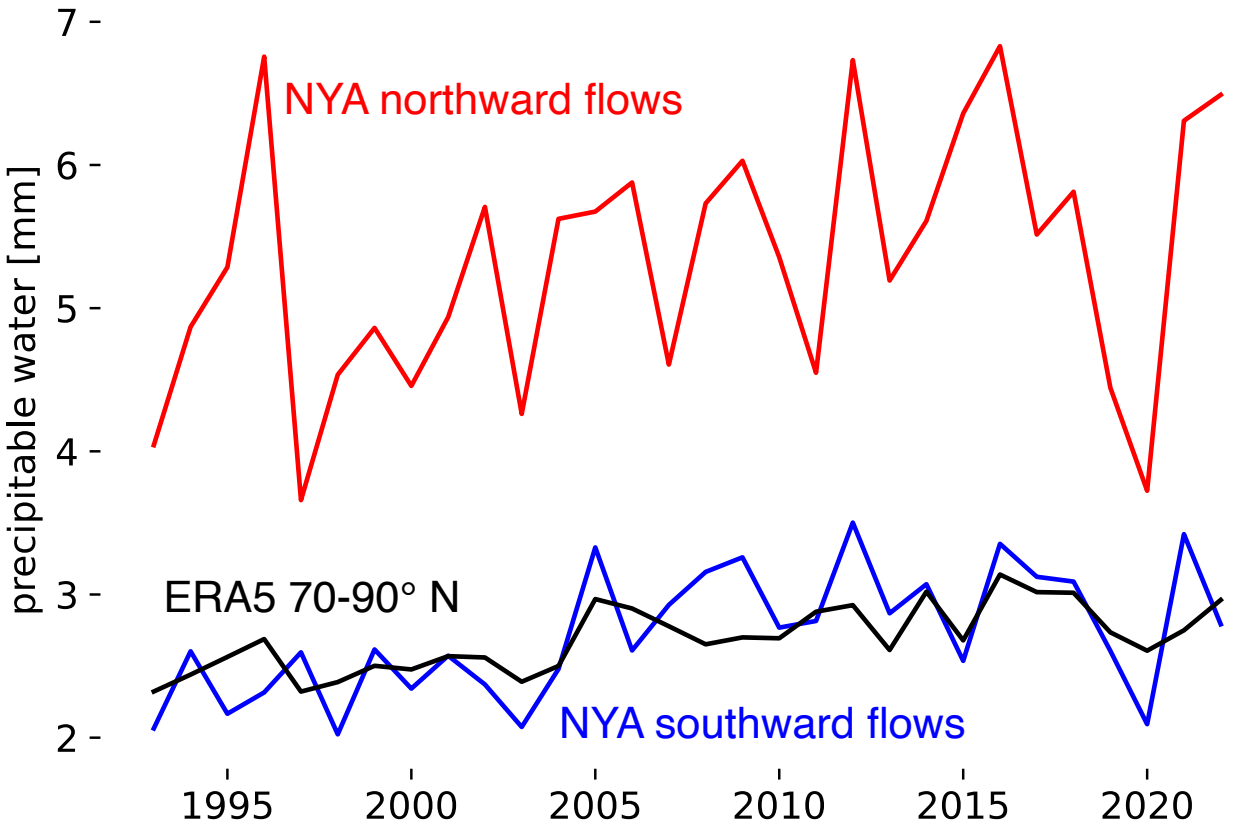

Figure 17. ~~a)~~ DJFM ~~annual~~ yearly mean of precipitable water observed over Ny Ålesund for southwesterly (red~~, upper line~~) and northerly flows (blue~~, lower line~~), and ~~b)~~ DJFM ~~annual~~ yearly mean of precipitable water in ERA5 averaged over the polar cap north of 70 $^{\circ}N$ (black line).

As reported from other Arctic wintertime observations (Stramler et al., 2011), the distribution of surface net longwave radiation in Ny Ålesund is bimodal, representing the clear and cloudy states of the Arctic winter boundary layer. The cloudy state is dominant during south-westerly advection, i.e. for air masses originating over open ocean, and the clear state is dominant for air masses arriving from the sea-ice covered Arctic ocean to the North of Svalbard. This bimodal distribution is not represented in all models. Two models substantially overestimate the downward longwave radiation out of cold air masses (in one case despite substantial atmospheric cold biases in these air masses). We attribute this to the cloudy state with high-emissivity, usually liquid-containing clouds being too frequent at temperatures substantially lower than -10 °C, where the clear state prevails in observations.

Within and across models, the typical moisture content of the atmosphere is strongly correlated between the grid point closest to Ny-Ålesund and a section across Fram Strait for both southerly and northerly flows. This suggests that the long-

term radiosonde observations at Ny Ålesund capture much of the interannual variability in the properties of air masses passing through this crucial gateway between the Arctic and lower latitudes, and that evaluating climate models against the observations provides a meaningful picture of the representation of air-mass properties in models.

The moisture content of air masses arriving in Ny Ålesund from the North shows a shift towards moister conditions in the early 2000s, which is also evident in reanalysis data averaged over the polar cap. This shift has been attributed to changes in the atmospheric circulation (Dahlke and Maturilli, 2017). The year 2020, i.e. the winter during the MOSAiC expedition, stands out as particularly dry compared to other years after this shift to moister conditions.

We conclude that the near-surface cold bias that occurs in many climate models over the Arctic ocean in winter is only partly reflected in free tropospheric temperatures ~~, and~~ in models with a modest overall cold bias. In all analysed models, the cold bias is much more pronounced for the coldest than for median temperatures. We recommend to further investigate the mechanism behind the frequent occurrence of strong subsaturation and its lack in coarse-resolution models. Finally, our results show that sub-sampling observations from stations around the Arctic ocean based on wind direction can help to detect trends over the central Arctic ocean, which is lacking long-term in-situ records.

Data availability. The Ny-Ålesund homogenized radiosonde data record is available at the PANGAEA data repository, for years 1993 to 2014 at https://doi.pangaea.de/10.1594/PANGAEA.845373 (Maturilli and Kayser, 2016), for years 2015 and 2016 at https://doi.pangaea.de/10.1594/PANG (Maturilli and Kayser, 2017b) and for years 2017 to 2022 at https://doi.org/10.1594/PANGAEA.961203 (Maturilli and Dünschede, 2023). The Ny-Ålesund surface radiation data at are freely available upon registration with the Baseline Surface Radiation Network (www.bsrn.awi.de), for years 1992 to 2013 at https://doi.pangaea.de/10.1594/PANGAEA.150000 (Maturilli et al., 2014) and for years 2006 to 2022 at https://doi.pangaea.de/10 (Maturilli, 2020).

The digital elevation model of Svalbard is provided by the Norwegian Polar Institute (Norwegian Polar Institute, 2014).

ICON output used in this paper will be archived and made publicly available at DKRZ upon publication.

CMIP6 data can be accessed through the ESGF system (Danek et al., 2020; NASA Goddard Institute for Space Studies (NASA/GISS), 2018; Yukimoto et al., 2019b; Jungclaus et al., 2019; Neubauer et al., 2019b; Tatebe and Watanabe, 2018; Ridley et al., 2019).

5

4.1

Author contributions. FP analyzed the data, produced the Figures and wrote the manuscript with input from MM and AKN. MM compiled the observational data and acquired funding. AKN produced ICON model runs. All authors contributed to the interpretation of results.

Competing interests. The authors declare no competing interests.

Acknowledgements. We thank Rune Graversen and one anonymous reviewer for their thorough review and useful suggestions, and the reviewers of the initial version of this paper for their critical remarks that have considerably improved the present manuscript. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003826 via project CRiceS (Climate Relevant interactions and feedbacks: the key role of sea ice and Snow in the polar and global climate system). We thank the staff of AWIPEV Research Base in Ny-Ålesund for reliably launching radiosondes over the years. FP acknowledges funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme under grant agreement No 101076205. A.K.N. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2037 'CLICCS—Climate, Climatic Change, and Society'—Project Number 390683824. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF.

References

Ali, S. M. and Pithan, F.: Following moist intrusions into the Arctic using SHEBA observations in a Lagrangian perspective, Quarterly Journal of the Royal Meteorological Society, 146, 3522–3533, 2020.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operationl convective-scale numerical weather prediction with the COSMO model: Description and sensitivities, Mon. Weather Rev., 139, 3887–3905, https://doi.org/10.1175/MWR-D-10-05013.1, 2011.

Beheng, K.: Numerical study on the combined action of droplet coagulation, ice particle riming and the splintering process concerning maritime cumuli, Contrib. Atmos. Phys.;(Germany, Federal Republic of), 55, 1982.

Bergeron, T.: Über die dreidimensional verknüpfende Wetteranalyse Erster Teil. Prinzipielle Einführung in das Problem der Luftmassen- und Frontenbil- dung., vol. 5 of *Geofys. Publ.*, Cammermeyers boghandel, 1928.

Boeke, R. C., Taylor, P. C., and Sejas, S. A.: On the nature of the Arctic's positive lapse-rate feedback, Geophysical Research Letters, 48, e2020GL091 109, 2021.

Bonan, D. B., Feldl, N., Siler, N., Kay, J. E., Armour, K. C., Eisenman, I., and Roe, G. H.: The influence of climate feedbacks on regional hydrological changes under global warming, Geophysical Research Letters, 51, e2023GL106 648, 2024.

Cesana, G., Kay, J., Chepfer, H., English, J., and De Boer, G.: Ubiquitous low-level liquid-containing Arctic clouds: New observations and climate model constraints from CALIPSO-GOCCP, Geophysical Research Letters, 39, 2012.

Cesana, G. V., Khadir, T., Chepfer, H., and Chiriaco, M.: Southern ocean solar reflection biases in CMIP6 models linked to cloud phase and vertical structure representations, Geophysical Research Letters, 49, e2022GL099 777, 2022.

Cronin, T. W. and Tziperman, E.: Low clouds suppress Arctic air formation and amplify high-latitude continental winter warming, Proceedings of the National Academy of Sciences, 112, 11 490–11 495, 2015.

Curry, J.: On the formation of continental polar air, Journal of the Atmospheric Sciences, 40, 2278–2292, 1983.

Dahlke, S. and Maturilli, M.: Contribution of atmospheric advection to the amplified winter warming in the Arctic North Atlantic region, Advances in Meteorology, 2017, Article ID 4928 620, https://doi.org/10.1155/2017/4928620, 2017.

Danek, C., Shi, X., Stepanek, C., Yang, H., Barbi, D., Hegewald, J., and Lohmann, G.: AWI AWI-ESM1.1LR model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.9328, 2020.

Davy, R. and Outten, S.: The Arctic surface climate in CMIP6: status and developments since CMIP5, Journal of Climate, 33, 8047–8068, 2020.

Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E., Denn, F. M., Duprat, T., Fukuda, M., Grobe, H., et al.: Baseline Surface Radiation Network (BSRN): structure and data description (1992–2017), Earth System Science Data, 10, 1491–1501, 2018.

Duffey, A., Mallett, R., Dutch, V. R., Steckling, J., Hermant, A., Day, J., and Pithan, F.: Stability of the Arctic Winter Atmospheric Boundary Layer over Sea Ice in CMIP6 Models, https://doi.org/10.22541/essoar.171405347.78397213/v1, manuscript submitted for publication, 2024.

Durre, I., Yin, X., Vose, R. S., Applequist, S., and Arnfield, J.: Enhancing the data coverage in the integrated global radiosonde archive, Journal of Atmospheric and Oceanic Technology, 35, 1753–1770, 2018.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, 2016.

Findeisen, W.: Colloidal meteorological processes in the formation of atmospheric precipitation, Meteorol. Z., 55, 121–133, 1938.

Graham, R. M., Hudson, S. R., and Maturilli, M.: Improved performance of ERA5 in Arctic gateway relative to four global atmospheric reanalyses, Geophysical Research Letters, 46, 6138–6147, 2019.

Graversen, R. G. and Burtu, M.: Arctic amplification enhanced by latent energy transport of atmospheric planetary waves, Quarterly Journal of the Royal Meteorological Society, 142, 2046–2054, 2016.

Hande, L., Engler, C., Hoose, C., and Tegen, I.: Seasonal variability of Saharan desert dust and ice nucleating particles over Europe, Atmospheric Chemistry and Physics, 15, 4389–4397, 2015.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., Bao, J., Bastin, S., Behravesh, M., Bergemann, M., et al.: ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales, Geoscientific Model Development, 16, 779–811, 2023.

Intrieri, J., Fairall, C., Shupe, M., Persson, P., Andreas, E., Guest, P., and Moritz, R.: An annual cycle of Arctic surface cloud forcing at SHEBA, Journal of Geophysical Research: Oceans, 107, SHE–13, 2002.

Jeffery, C. and Austin, P.: Homogeneous nucleation of supercooled water: Results from a new equation of state, Journal of Geophysical Research: Atmospheres, 102, 25 269–25 279, 1997.

Juckes, M., Taylor, K. E., Durack, P. J., Lawrence, B., Mizielinski, M. S., Pamment, A., Peterschmitt, J.-Y., Rixen, M., and Sénési, S.: The CMIP6 data request (DREQ, version 01.00. 31), Geoscientific Model Development, 13, 201–224, 2020.

Jungclaus, J., Bittner, M., Wieners, K.-H., Wachsmann, F., Schupfner, M., Legutke, S., Giorgetta, M., Reick, C., Gayler, V., Haak, H., de Vrese, P., Raddatz, T., Esch, M., Mauritsen, T., von Storch, J.-S., Behrens, J., Brovkin, V., Claussen, M., Crueger, T., Fast, I., Fiedler, S., Hagemann, S., Hohenegger, C., Jahns, T., Kloster, S., Kinne, S., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Müller, W., Nabel, J., Notz, D., Peters-von Gehlen, K., Pincus, R., Pohlmann, H., Pongratz, J., Rast, S., Schmidt, H., Schnur, R., Schulzweida, U., Six, K., Stevens, B., Voigt, A., and Roeckner, E.: MPI-M MPI-ESM1.2-HR model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.6594, 2019.

Kärcher, B., Hendricks, J., and Lohmann, U.: Physically based parameterization of cirrus cloud formation for use in global atmospheric models, Journal of Geophysical Research: Atmospheres, 111, 2006.

Karlsson, J. and Svensson, G.: The simulation of Arctic clouds and their influence on the winter surface temperature in present-day climate in the CMIP3 multi-model dataset, Climate Dynamics, 36, 623–635, 2011.

Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell, G. L., Ackerman, A. S., Aleinov, I., Bauer, M., Bleck, R., et al.: GISS-E2. 1: Configurations and climatology, Journal of Advances in Modeling Earth Systems, 12, e2019MS002 025, 2020.

Komurcu, M., Storelvmo, T., Tan, I., Lohmann, U., Yun, Y., Penner, J. E., Wang, Y., Liu, X., and Takemura, T.: Intercomparison of the cloud water phase among global climate models, Journal of Geophysical Research: Atmospheres, 119, 3372–3400, 2014.

Lawrence, H., Bormann, N., Sandu, I., Day, J., Farnan, J., and Bauer, P.: Use and impact of Arctic observations in the ECMWF Numerical Weather Prediction system, Quarterly Journal of the Royal Meteorological Society, 145, 3432–3454, 2019.

Lawrence, Z. D., Perlwitz, J., Butler, A. H., Manney, G. L., Newman, P. A., Lee, S. H., and Nash, E. R.: The remarkably strong Arctic stratospheric polar vortex of winter 2020: Links to record-breaking Arctic oscillation and ozone loss, Journal of Geophysical Research: Atmospheres, 125, e2020JD033 271, 2020.

Linke, O., Quaas, J., Baumer, F., Becker, S., Chylik, J., Dahlke, S., Ehrlich, A., Handorf, D., Jacobi, C., Kalesse-Los, H., Lelli, L., Mehrdad, S., Neggers, R. A. J., Riebold, J., Saavedra Garfias, P., Schnierstein, N., Shupe, M. D., Smith, C., Spreen, G., Verneuil, B., Vinjamuri,

K. S., Vountas, M., and Wendisch, M.: Constraints on simulated past Arctic amplification and lapse rate feedback from observations, Atmospheric Chemistry and Physics, 23, 9963–9992, https://doi.org/10.5194/acp-23-9963-2023, 2023.

Manabe, S. and Wetherald, R. T.: The effects of doubling the CO 2 concentration on the climate of a general circulation model, Journal of Atmospheric Sciences, 32, 3–15, 1975.

Maturilli, M.: Basic and other measurements of radiation at station Ny-Ålesund (2006-05 et seq), https://doi.org/10.1594/PANGAEA.914927, 2020.

Maturilli, M. and Dünschede, E.: Homogenized radiosonde record at station Ny-Ålesund, Spitsbergen, 2017-2022., PANGAEA, https://doi.org/10.1594/PANGAEA.961203, 2023.

Maturilli, M. and Kayser, M.: Homogenized radiosonde record at station Ny-Ålesund, Spitsbergen, 1993-2014, https://doi.org/10.1594/PANGAEA.845373, 2016.

Maturilli, M. and Kayser, M.: Arctic warming, moisture increase and circulation changes observed in the Ny-Ålesund homogenized radiosonde record, Theoretical and Applied Climatology, 130, 1–17, 2017a.

Maturilli, M. and Kayser, M.: Homogenized radiosonde record at station Ny-Ålesund, Spitsbergen, 2015-2016, https://doi.org/10.1594/PANGAEA.875196, 2017b.

Maturilli, M., Herber, A., and König-Langlo, G.: Basic and other measurements of radiation from the Baseline Surface Radiation Network (BSRN) Station Ny-Ålesund in the years 1992 to 2013, reference list of 253 datasets, https://doi.org/10.1594/PANGAEA.150000, 2014.

Maturilli, M., Herber, A., and König-Langlo, G.: Surface radiation climatology for Ny-Ålesund, Svalbard (78.9 N), basic observations for trend detection, Theoretical and Applied Climatology, 120, 331–339, 2015.

Mayer, M., Tietsche, S., Haimberger, L., Tsubouchi, T., Mayer, J., and Zuo, H.: An improved estimate of the coupled Arctic energy budget, Journal of Climate, 32, 7915–7934, 2019.

Medeiros, B., Deser, C., Tomas, R. A., and Kay, J. E.: Arctic inversion strength in climate models, Journal of Climate, 24, 4733–4740, 2011.

NASA Goddard Institute for Space Studies (NASA/GISS): NASA-GISS GISS-E2.1G model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.7127, 2018.

Naumann, A. K., Esch, M., and Stevens, B.: How the representation of microphysical processes affects tropical condensate in a global storm-resolving model, EGUsphere, 2024, 1–19, https://doi.org/10.5194/egusphere-2024-2268, 2024.

Neubauer, D., Ferrachat, S., Siegenthaler-Le Drian, C., Stier, P., Partridge, D. G., Tegen, I., Bey, I., Stanelle, T., Kokkola, H., and Lohmann, U.: The global aerosol–climate model ECHAM6. 3–HAM2. 3–Part 2: Cloud evaluation, aerosol radiative forcing, and climate sensitivity, Geoscientific Model Development, 12, 3609–3639, 2019a.

Neubauer, D., Ferrachat, S., Siegenthaler-Le Drian, C., Stoll, J., Folini, D. S., Tegen, I., Wieners, K.-H., Mauritsen, T., Stemmler, I., Barthel, S., Bey, I., Daskalakis, N., Heinold, B., Kokkola, H., Partridge, D., Rast, S., Schmidt, H., Schutgens, N., Stanelle, T., Stier, P., Watson-Parris, D., and Lohmann, U.: HAMMOZ-Consortium MPI-ESM1.2-HAM model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.5016, 2019b.

Norwegian Polar Institute: Terrengmodell Svalbard (S0 Terrengmodell), https://doi.org/10.21334/npolar.2014.dce53a47, 2014.

Nygård, T., Naakka, T., and Vihma, T.: Horizontal moisture transport dominates the regional moistening patterns in the Arctic, Journal of Climate, 33, 6793–6807, 2020.

Papritz, L. and Sodemann, H.: Characterizing the local and intense water cycle during a cold air outbreak in the Nordic seas, Monthly Weather Review, 146, 3567–3588, 2018.

Pithan, F. and Jung, T.: Arctic amplification of precipitation changes—The energy hypothesis, Geophysical Research Letters, 48, e2021GL094 977, 2021.

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

Pithan, F., Medeiros, B., and Mauritsen, T.: Mixed-phase clouds cause climate model biases in Arctic wintertime temperature inversions, Climate dynamics, 43, 289–303, 2014.

Pithan, F., Svensson, G., Caballero, R., Chechin, D., Cronin, T. W., Ekman, A. M., Neggers, R., Shupe, M. D., Solomon, A., Tjernström, M., et al.: Role of air-mass transformations in exchange between the Arctic and mid-latitudes, Nature Geoscience, 11, 805–812, 2018.

Quaas, J.: Evaluating the "critical relative humidity" as a measure of subgrid-scale variability of humidity in general circulation model cloud cover parameterizations using satellite data, Journal of Geophysical Research: Atmospheres, 117, 2012.

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., and Andrews, T.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.6109, 2019.

Sandu, I., van Niekerk, A., Shepherd, T. G., Vosper, S. B., Zadra, A., Bacmeister, J., Beljaars, A., Brown, A. R., Dörnbrack, A., McFarlane, N., et al.: Impacts of orography on large-scale atmospheric circulation, npj Climate and Atmospheric Science, 2, 10, 2019.

Schön, M., Suomi, I., Altstädter, B., van Kesteren, B., zum Berge, K., Platis, A., Wehner, B., Lampert, A., and Bange, J.: Case studies of the wind field around Ny-Ålesund, Svalbard, using unmanned aircraft, Polar Research, 41, 2022.

Seifert, A. and Beheng, K.: A two-moment cloud microphysics parameterization for mixed-phase clouds. Part I: Model Description, Meteorol. Atmos. Phys., 92, 45–66, 2006.

Serreze, M. C., Kahl, J. D., and Schnell, R. C.: Low-level temperature inversions of the Eurasian Arctic and comparisons with Soviet drifting station data, Journal of Climate, 5, 615–629, 1992.

Shestakova, A. A., Chechin, D. G., Lüpkes, C., Hartmann, J., and Maturilli, M.: The foehn effect during easterly flow over Svalbard, Atmospheric Chemistry and Physics, 22, 1529–1548, 2022.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., et al.: Atmospheric component of the MPI-M Earth system model: ECHAM6, Journal of Advances in Modeling Earth Systems, 5, 146–172, 2013.

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., Düben, P., Judt, F., Khairoutdinov, M., Klocke, D., et al.: DYAMOND: The DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains, 6, 61, 2019a.

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., Düben, P., Judt, F., Khairoutdinov, M., Klocke, D., et al.: DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains, Progress in Earth and Planetary Science, 6, 1–17, 2019b.

Storelvmo, T. and Tan, I.: The Wegener–Bergeron–Findeisen process—Its discovery and vital importance for weather and climate, Meteorologische Zeitschrift, 24, 455–461, 2015.

Stramler, K., Del Genio, A. D., and Rossow, W. B.: Synoptically driven Arctic winter states, Journal of Climate, 24, 1747–1762, 2011.

Sundqvist, H.: A parameterization scheme for non-convective condensation including prediction of cloud water content, Quarterly Journal of the Royal Meteorological Society, 104, 677–690, 1978.

Svensson, G. and Karlsson, J.: On the Arctic wintertime climate in global climate models, Journal of climate, 24, 5757–5771, 2011.

Tan, I., Barahona, D., and Coopman, Q.: Potential link between ice nucleation and climate model spread in Arctic amplification, Geophysical Research Letters, 49, e2021GL097 373, 2022.

Tatebe, H. and Watanabe, M.: MIROC MIROC6 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.5603, 2018.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., et al.: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6, Geoscientific Model Development, 12, 2727–2765, 2019.

Tompkins, A. M., Gierens, K., and Rädel, G.: Ice supersaturation in the ECMWF integrated forecast system, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 133, 53–63, 2007.

Verlinde, J., Harrington, J. Y., McFarquhar, G., Yannuzzi, V., Avramov, A., Greenberg, S., Johnson, N., Zhang, G., Poellot, M., Mather, J. H., et al.: The mixed-phase Arctic cloud experiment, Bulletin of the American Meteorological Society, 88, 205–222, 2007.

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., et al.: The Met Office Unified Model global atmosphere 7.0/7.1 and JULES global land 7.0 configurations, Geoscientific Model Development, 12, 1909–1963, 2019.

Wegener, A.: Thermodynamik der atmosphäre, JA Barth, 1911.

Wexler, H.: Cooling in the lower atmosphere and the structure of polar continental air, Monthly Weather Review, 64, 122–136, 1936.

Winkelbauer, S., Mayer, M., and Haimberger, L.: Validation of key Arctic energy and water budget components in CMIP6, Climate Dynamics, pp. 1–36, 2024.

Woods, C. and Caballero, R.: The role of moist intrusions in winter Arctic warming and sea ice decline, Journal of Climate, 29, 4473–4485, 2016.

Woods, C., Caballero, R., and Svensson, G.: Large-scale circulation associated with moisture intrusions into the Arctic during winter, Geophysical Research Letters, 40, 4717–4721, 2013.

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., et al.: The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2. 0: Description and basic evaluation of the physical component, Journal of the Meteorological Society of Japan. Ser. II, 97, 931–965, 2019a.

Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yoshimura, H., Shindo, E., Mizuta, R., Ishii, M., Obata, A., and Adachi, Y.: MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.6842, 2019b.