JGR Atmospheres

RESEARCH ARTICLE
10.1029/2019JD031783

Confronting Arctic Troposphere, Clouds, and Surface Energy Budget Representations in Regional Climate Models With Observations

Joseph Sedlar[1,2], Michael Tjernström[3], Annette Rinke[4], Andrew Orr[5], John Cassano[1,6], Xavier Fettweis[7], Günther Heinemann[8], Mark Seefeldt[1], Amy Solomon[1,9], Heidrun Matthes[4], Tony Phillips[5], and Stuart Webster[10]

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA, [2]Global Monitoring Division, NOAA Earth System Research Laboratory, Boulder, CO, USA, [3]Department of Meteorology and Bert Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden, [4]Alfred Wegener Institute, Helmholtz Centre for Polar and Marie Research, Potsdam, Germany, [5]British Antarctic Survey, National Environmental Research Council, Cambridge, UK, [6]Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO, USA, [7]Department of Geography, University of Liège, Liège, Belgium, [8]Department of Environmental Meteorology, University of Trier, Trier, Germany, [9]Physical Sciences Division, NOAA Earth System Research Laboratory, Boulder, CO, USA, [10]Met Office, Exeter, UK

Key Points:
- Regional climate model simulations of the Arctic summer and autumn atmosphere are evaluated against detailed field campaign observations during 2014
- Modeled variability in the simulation of low-level liquid containing clouds led to critical errors in surface energy budget terms
- Clouds, surface energy fluxes, and lower tropospheric stability are interdependent, and the models have difficulty representing these critical connections

Supporting Information:
- Supporting Information S1

Correspondence to:
J. Sedlar,
joseph.sedlar@colorado.edu

Citation:
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*, *124*. https://doi.org/10.1029/2019JD031783

Received 7 OCT 2019
Accepted 28 FEB 2020
Accepted article online 12 MAR 2020

Abstract A coordinated regional climate model (RCM) evaluation and intercomparison project based on observations from a July–October 2014 trans-Arctic Ocean field experiment (ACSE-Arctic Clouds during Summer Experiment) is presented. Six state-of-the-art RCMs were constrained with common reanalysis lateral boundary forcing and upper troposphere nudging techniques to explore how the RCMs represented the evolution of the surface energy budget (SEB) components and their relation to cloud properties. We find that the main reasons for the modeled differences in the SEB components are a direct consequence of the RCM treatment of cloud and cloud-radiative interactions. The RCMs could be separated into groups by their overestimation or underestimation of cloud liquid. While radiative and turbulent heat flux errors were relatively large, they often invoke compensating errors. In addition, having the surface sea-ice concentrations constrained by the reanalysis or satellite observations limited how errors in the modeled radiative fluxes could affect the SEB and ultimately the surface evolution and its coupling with lower tropospheric mixing and cloud properties. Many of these results are consistent with RCM biases reported in studies over a decade ago. One of the six models was a fully coupled ocean-ice-atmosphere model. Despite the biases in overestimating cloud liquid, and associated SEB errors due to too optically thick clouds, its simulations were useful in understanding how the fully coupled system is forced by, and responds to, the SEB evolution. Moving forward, we suggest that development of RCM studies need to consider the fully coupled climate system.

1. Introduction

Regional climate models (RCMs) provide a valuable tool to improve our understanding of the Arctic climate system, a region that continues to experience unprecedented climate change (e.g., AMAP, 2017; Vaughan et al., 2013). Since the spatial and temporal coverage by the observational network is sparse in the Arctic, RCMs are a means to provide climatological information at a high resolution with full spatial coverage (e.g., Gutjahr & Heinemann, 2018; Kohnemann et al., 2017). Thus, it is of great importance to evaluate the current state of the art of RCM modeling in the Arctic. The large-scale climate of RCMs is generally constrained by global atmospheric models, either from reanalyses or climate models (GCMs), driving the lateral boundaries of the RCM (e.g., Giorgi et al., 2009). Using a reanalysis, the constrained boundaries act to minimize large-scale dynamic and thermodynamic drift from reality, permitting the RCM to simulate the full dynamical system within a smaller domain at relatively high horizontal and vertical resolutions, generally much higher than current GCMs. If the RCMs are also be nudged to the large-scale reanalysis in the upper troposphere (i.e., Akperov et al., 2018; Berg et al., 2013), such an exercise can provide insight into model soundness and model shortcomings (Rinke et al., 2000) by comparison to observations. This can guide future development of model dynamics, thermodynamics, and descriptions of small-scale processes

(parameterizations) that are subgrid scale for tomorrow's global climate models, eventually when running on similar resolutions as the current RCMs (Cassano et al., 2017).

Evaluation studies of RCM simulations across the Arctic have proved to be a rich exercise. Rinke et al. (2000) performed multimodel RCM simulations of the Arctic region poleward of 65°N for one winter and one summer month in 1990 with two RCMs. They found substantial differences in the surface energy budget (SEB) components that could be traced back to biases in the clouds and cloud properties, as well as to differences in the treatment of sea-ice extent within a model grid point. Biases in the SEB translated to errors in the surface temperature, further biasing the near-surface temperature structure and stratification (Rinke et al., 2000). Systematic errors emerge because of an imperfect understanding of the coupled system, including the ocean, land, cryosphere, and atmosphere. The generality of physical parameterizations further poses a problem for the data-sparse Arctic, where observations of small-scale processes, often distinct from other regions on Earth (Rinke et al., 2000), are few. Still, these observations are needed to improve foundational process understanding.

The Arctic Regional Climate Model Intercomparison Project (ARCMIP; Curry & Lynch, 2002) coordinated common surface and lateral boundary forcings for eight RCMs to simulate the annual evolution for the 1997–1998 SHEBA ice-drift atmospheric and SEB observations (Uttal et al., 2002). Despite reasonable simulations of the large-scale atmospheric environment (Rinke et al., 2006; Tjernström et al., 2005), near-surface state variables contained obvious biases that were directly related to biases in the near-surface radiative and turbulent heat fluxes (Tjernström et al., 2005, 2008; Wyser et al., 2008). The forcing of the Arctic SEB, especially the radiative fluxes, is critically dependent upon the cloud macrophysical and microphysical properties (e.g., Intrieri et al., 2002; Klein et al., 2009; Persson, 2012; Sedlar et al., 2011; Shupe & Intrieri, 2004; Tjernström et al., 2008). Wyser et al. (2008) reported that even though the simulations poorly captured the observed instantaneous clouds, the correlations between surface radiation in the ARCMIP simulations and the SHEBA observations were high. Errors in the amount of total cloud water/ice mixing ratio, particularly the partitioning between liquid and ice, largely contributed to the modeled biases in surface radiation (Inoue et al., 2006; Tjernström et al., 2005, 2008). Tjernström et al. (2008) further concluded that even if the partitioning of cloud phase were accurate, the prescribed manner by which the modeled microphysics interacted with the radiative transfer may exacerbate the surface radiation biases.

Despite relatively weak turbulent heat fluxes over summer Arctic sea ice, the observed sum of these fluxes is often of the same order of magnitude as the SEB residual (Sedlar et al., 2011; Tjernström et al., 2005); an accurate representation of the turbulent heat fluxes is therefore critical, especially if RCMs are coupled to sea-ice/ocean models. At SHEBA, the absolute magnitude of the modeled sensible and latent heat fluxes (SHF and LHF) was not inconsistent with the observed turbulent fluxes (Tjernström et al., 2005). However, instantaneous errors in modeled minus observed SHF and LHF were frequently as large as the observed fluxes. These fluxes are dependent upon the surface characteristics, winds, and the lower tropospheric stratification, which is very closely linked to low-level Arctic mixed-phase clouds (Brooks et al., 2017; Persson et al., 2017; Sedlar et al., 2011; Sedlar & Shupe, 2014; Shupe et al., 2008, 2013; Sotiropoulou et al., 2014; Tjernström et al., 2012). Known biases in modeled clouds can therefore directly influence biases in modeled turbulent heat fluxes.

Given the advancement in modeling in recent years, such as increased grid resolution and improved understanding of key physical processes and their representations in RCMs, it is timely to assess the ability of RCMs to simulate the linkages between bulk cloud properties and the SEB over the Arctic. Following the legacy from ARCMIP, where RCMs simulated the drifting SHEBA observatory, this study describes a coordinated RCM intercomparison where models simulated an Arctic field campaign, the Arctic Clouds in Summer Experiment (ACSE), from early July into early October 2014. ACSE was deployed on the Swedish research icebreaker Oden during a polar transit that navigated the eastern Arctic, from northern Europe, following northern Eurasia to Alaska, and back to Europe (Figure 1; Sotiropoulou et al., 2016; Tjernström et al., 2015, 2019). A suite of six RCMs with established Arctic domains from the CORDEX (Coordinated Regional Downscaling Experiment, Giorgi et al., 2009) project contributed to this intercomparison, running the simulations and extracting output at the moving ACSE location.

This study provides insight into a range of model-observation biases for current state-of-the-art RCM representations of clouds, radiation, and surface energy fluxes in the contemporary Arctic climate. Results and

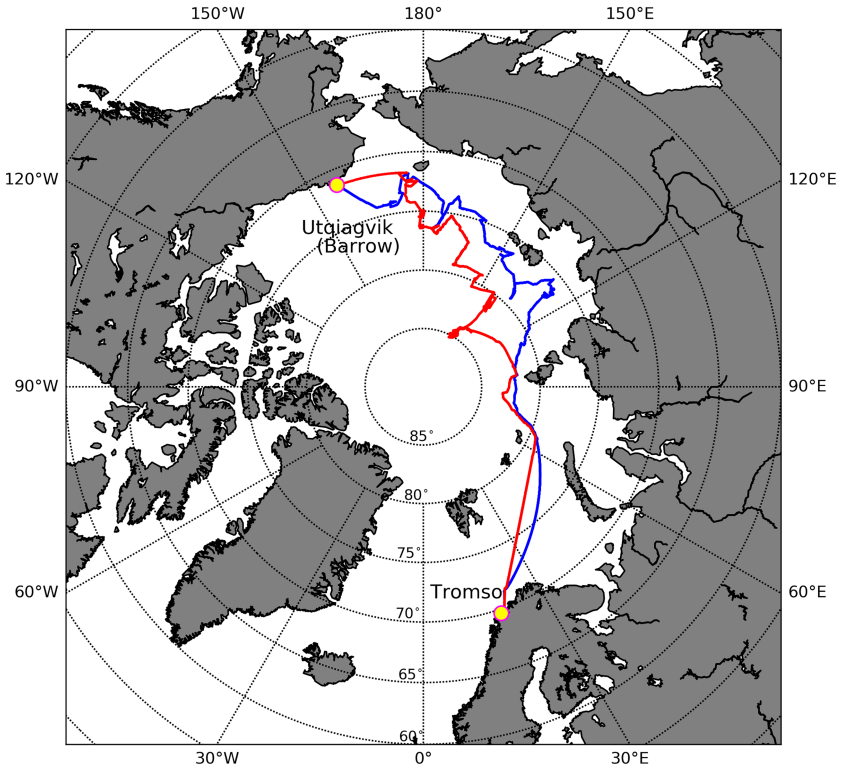

Figure 1. Arctic Clouds in Summer Experiment (ACSE) cruise track depicting: (blue) the outbound transit from Tromsö, Norway, to Utqiagvik (Barrow), Alaska, from July 5 to August 18, and (red) the inbound transit back from 21 August to 5 October 2014.

conclusions from the intercomparisons will be useful for modelers in order to explore potential parameterization development and improvement strategies ahead of the completion of the yearlong international Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) field experiment, for which coordinated RCM experiments are anticipated. While we attempt to isolate causes for discrepancy and spread in cloud properties and surface energy fluxes among the models, it is beyond the scope of this study to delve into each individual model's physical parameterizations to pinpoint root causes. Instead, we aim to provide guidance to the respective modeling centers identifying where models adequately reproduce physical processes and when critical shortcomings are found. This work is a contribution to the Polar CORDEX project (e.g., Giorgi et al., 2009) and comes nearly two decades after the ARCMIP simulations, hence presenting a unique opportunity to investigate whether increased model grid resolution or enhancements to physical parameterizations since ARCMIP leads to improvements in Arctic clouds and radiation.

2. Models, Experimental Design, and Observations

2.1. Regional Climate Models

Table 1 provides a summary of the five atmospheric RCMs and the one coupled RCM. Each model provided at least one full simulation, while CCLM, HIRHAM5, and CAFS also contributed additional explorative runs. Two CCLM simulations were forced with different lateral boundary forcing: ERA-Interim (Dee et al., 2011; hereafter cclmi) and ERA5 (Hersbach et al., 2018; hereafter cclm5); CCLM uses ERA5 data with a temporal resolution of 1 hr. Three HIRHAM5 simulations differed in cloud scheme and sea ice; hirham5_v1 and hirham5_v3 employed the Sundqvist et al. (1989) cloud scheme but used different prescribed sea-ice thickness, 2 and 1.5 m, respectively, while hirham5_v2 ran the Tompkins (2002) cloud scheme with prescribed 2-m ice thickness. CAFS was run in one baseline simulation (cafs_bsl), while cafs_ini featured

Journal of Geophysical Research: Atmospheres

10.1029/2019JD031783

Table 1
Regional Climate Model Forcing and Simulation Name Information, Vertical and Horizontal Resolution, Atmospheric Nudging Technique Employed, Sea-Ice Concentration Initialization, and Relevant References

Institution/country	Model name	Forcing data	Simulation name(s)	Resolution Vertical, horizontal	Large-scale dynamic constraint	Sea-ice concentration	Sea ice thickness	References
U. Trier/Germany	CCLM	ERA-Interim ERA5	cclmi cclm5	L60, 0.125° (~15 × 15 km)	Forecast mode (initialized daily)	Daily AMSR2 data	PIOMAS daily data (Zhang & Rothrock, 2003)	Schröder et al. (2011); Gutjahr and Heinemann (2018)
AWI/Germany	HIRHAM5	ERA-Interim	hirham5_v1 hirham5_v2 hirham5_v3	L40, 0.5° (~50 × 50 km)	Grid point nudging at all levels (T, U, V, Q)	Daily ERA-Interim	2 m or 1.5 m (v3)	Christensen et al. (2007); Sommerfeld et al. (2015); Klaus et al. (2016)
U. Colorado/USA	WRF	ERA-Interim	wrf_cu	L40, 25 × 25 km	Grid point nudging (T, U, V) top 20 levels	ERA-Interim	3 m	Powers et al. (2017) Skamarock et al. (2008)
NOAA-ESRL-CIRES/ USA	CAFS	ERA-Interim	cafs_bslcafs_ini	L42, ~10 × 10 km	Grid point nudging (T, U, V, Q) above 500 hPa	Coupled model	Coupled model	Intrieri et al., in preparation
UK Met Office/UK	Met Office Unified Model (vn11.1)	ERA-Interim	metum	L70, 0.225° (25 × 25 km)	Forecast mode (initialized twice daily, with output archived from T + 12 hr to T + 24 hr)	Twice daily ERA-Interim data	2 m	Orr et al. (2014) Walters et al. (2017)
U. Liege/Belgium	MAR3.9.6	ERA-Interim	mar_v3.9.6	L24, 25 × 25 km	Spectral (U, V, T for lower stratosphere)	ERA-Interim	0.5 m	Fettweis et al. (2017)

Table 2
Regional Climate Model Specifications, Including Number of Vertical Levels and Number of Levels Below 1,000 and 500 m, the Lowest Vertical Model Level Above the Surface, and the Parameterization Schemes

Model name	Simulation name	Forcing data	Vertical levels Total/below 1,000 m/500 m	Lowest model level	Surface layer scheme	PBL scheme	Cloud cover scheme/ radiation
CCLM	cclmi cclm5	ERA-Interim ERA5	60/18/12	10 m	Two-layer thermodynamic sea-ice model, Louis (1979) scheme, tile approach for fluxes	level-2.5 closure scheme (Mellor & Yamada, 1974)	Ritter and Geleyn (1992), Doms et al. (2011)
HIRHAM5	hirham5_v1 hirham5_v2 hirham5_v3	ERA-Interim	40/10/7	~10 m	Louis scheme (Louis, 1979)	Level 1.5 (TKE closure; Brinkop & Roeckner, 1995)	v1: Sundqvist et al. (1989) v2: Tompkins (2002) v3: Sundqvist et al. (1989)
WRF-CU	wrf_cu	ERA-Interim	40/10/8	~12 m AGL	Revised MM5 (Jiménez et al., 2012)	MYNN 2.5 (Nakanishi & Niino, 2006)	Morrison et al. (2009)
CAFS	cafs_bsl cafs_ini	ERA-Interim	42/10/8	12 m	Revised MM5 (Jiménez et al., 2012)	YSU (Hong et al., 2006)	Morrison et al. (2009) Ice nuclei concentration: cafs_bsl = 0.16 L^{-1} cafs_ini = 1.0 L^{-1}
MetUM	Metum	ERA-Interim	70/11/7	20 m	JULES scheme (Best et al., 2011)	Lock et al. (2000), using the sharpest stability function	Wilson and Ballard (1999), Wilson et al. (2008)
MARv3.9.6	mar_v3.9.6	ERA-Interim	24/12/11	2.5 m	SISVAT (Gallèe et al., 2001)	k-epsilon closure (Duynkerke, 1988)	Gallèe et al. (2001)

Abbreviation: PBL: planetary boundary layer.

increased ice nuclei (IN) concentration, from 0.16 to 1 L^{-1}, to test the sensitivity of cloud phase partitioning on IN concentration.

There is a wide range among models in horizontal and vertical grid resolutions (Table 1). Except for hirham5, horizontal resolutions were 25 km or better, at least twice as high as in ARCMIP (Rinke et al., 2006). The number of vertical levels increased, from 20–30 during ARCMIP, to 40–70 (mar_v3.9.6 is an exception with 24 levels) in the current evaluation. The largest resolution increase was typically in the lower troposphere (Table 2). With ARCMIP, three to four levels were below 500 m above ground level (AGL), here that increased about twofold. The lowest model level, from 2.5 to 20 m AGL, is also reduced form the 15–80 m range in ARCMIP (Tjernström et al., 2005). A list of the model surface layer, planetary boundary layer (PBL), and cloud/radiation schemes is presented in Table 2.

2.2. Model Experimental Design

All six models provided at least one simulation of the ACSE period (section 2.3) forced by ERA-Interim reanalysis at the lateral boundaries over the Arctic CORDEX domain (https://www.cordex.org/domains/region-11-arctic/). To minimize model drift due to the large, pan-Arctic domain, different models used different techniques to constrain the larger-scale dynamics of the RCMs (e.g., Giorgi et al., 2009): spectral or grid point nudging of large-scale state variables or frequent model reinitialization (Table 1).

Sea-ice cover fractions were prescribed from satellite information, either directly or passed through the reanalysis, except for in cafs. This was the only fully coupled ocean-ice-atmosphere model and therefore only included prescribed sea-ice concentration and thickness at the model initiation; cafs was then free to evolve its own surface properties as dictated by its own SEB. Models calculated the effective surface skin temperature with a tiled-averaging approach, using the open-water and sea-ice fractions prescribed. While sea-ice temperature was determined by each model's SEB evolution, sea-surface temperatures were updated daily or multiple times per day from reanalysis. Model time step was also specified individually for each RCM.

Variables were extracted hourly at the grid point nearest the ACSE location. Near-surface radiation, turbulent heat fluxes, and integrated cloud water paths were extracted as 1-hr averages. All remaining variables, including thermodynamic and cloud water content profiles, near-surface winds, surface temperature, sea-ice

concentration, and surface albedo, were instantaneous. Comparison of RCMs with observations occurred for 01:00 UTC on 5 July 2014 to 07:00 UTC on 3 October, at a frequency of 1 hr.

It is important to understand the reason for these constraints. To compare model results to observations that were taken at different locations along a time line, and to be able to asses and identify discrepancies due to model deficiencies, the RCMs must be controlled such that they describe the larger-scale meteorological conditions correctly. Due to the inherent nonlinearity of atmospheric dynamics, if not controlled this way, each free running model would develop its own weather development, all different and equally plausible but in contrast with what actually happened. It would be difficult if not impossible to analyze what differences between the models and the observations resulted from inadequate small-scale parameterizations and what would be due differences in large-scale weather. Herein lies one strength of using RCMs. By controlling the larger-scale dynamics, it becomes possible to compare model results directly to observations and to assess the adequacy of the models' process descriptions. The experimental design therefore rests on forcing the models to realistically describe larger-scale conditions while allowing them full freedom to develop systematic errors depending on the description of the small-scale processes, including all the nonlinear feedback between different processes.

2.3. ACSE Observations

Unlike SHEBA, ACSE was based onboard a navigating icebreaker (Figure 1) and only sampled during two seasons, from 5 July to 3 October in 2014. The sampling conditions during ACSE varied across a range of surface types, changing between open water and partial or complete sea-ice cover, with and without melt ponds. Ice-covered and ice-free regions were sampled in frequency roughly the same during both seasons (Sotiropoulou et al., 2016). ACSE was planned to observe two important seasons, summer melt and autumn freeze, and the transition between them; Sotiropoulou et al. (2016) concluded that the seasonal transition was observed.

Detailed descriptions of the instrumentation and measurements are provided in Sotiropoulou et al. (2016) and Tjernström et al. (2019). We examine only a subset of the available observations. Atmospheric profiling was made by 6-hourly radiosoundings with temperature and pressure uncertainties documented to 0.5 °C and 1 hPa, respectively. Relative humidity measurements are more complicated, with an accuracy of 5% (Vaisala RS92 User Manual); specific humidity was calculated from Clausius-Clapeyron. The radiosoundings were assimilated in the reanalyses.

A weather station was deployed on the upper-most deck (~25 m AGL) of the icebreaker, combining measurements of pressure, temperature, and relative humidity from a Vaisala PTU300, with horizontal wind speed and direction from a heated Gill WindSonicM. The 25-m level provided the least interference from the ship's superstructure for downwelling radiative fluxes. An Eppley Precision Spectral Pyranometer and a Precision Infrared Radiometer measured the broadband shortwave downwelling (SWD) and longwave downwelling (LWD) fluxes. Extensive quality controls were applied to flag, and if possible correct, radiation measurements, especially for the shortwave. Solar azimuth and zenith angles were computed to identify times when direct solar irradiance may have been influenced by shading from physical structures on the upper deck. Upwelling radiation measurements were not made during ACSE; hence, albedo was not measured. Heitronics KT15 infrared radiation thermometers viewing the surface on both sides of the ship provided surface skin temperature with 0.5 °C accuracy. We used the Stefan Boltzmann relationship combining skin temperature and an assumption of surface infrared emissivity (0.99; e.g., Andreas et al., 2010) to estimate the surface upwelling longwave flux.

Cloud properties are examined mainly using a bulk methodology. Retrievals from a dual-channel microwave radiometer provided cloud liquid water path (LWP) and precipitable water (PW) with uncertainties of 25 and 0.1 kg/m^2, respectively (Westwater et al., 2001). Column-integrated ice water path (IWP) was retrieved from a combination of lidar, cloud radar, and thermodynamic profiles using the CloudNet retrieval algorithm (Illingworth et al., 2007; Achtert et al., in preparation). The uncertainty inherent in such retrievals can be as large as a factor of two (Shupe et al., 2004) and must be considered with caution. We have also used the vertical retrievals of liquid and ice water contents (LWC and IWC) to explore the vertical partitioning of hydrometeors.

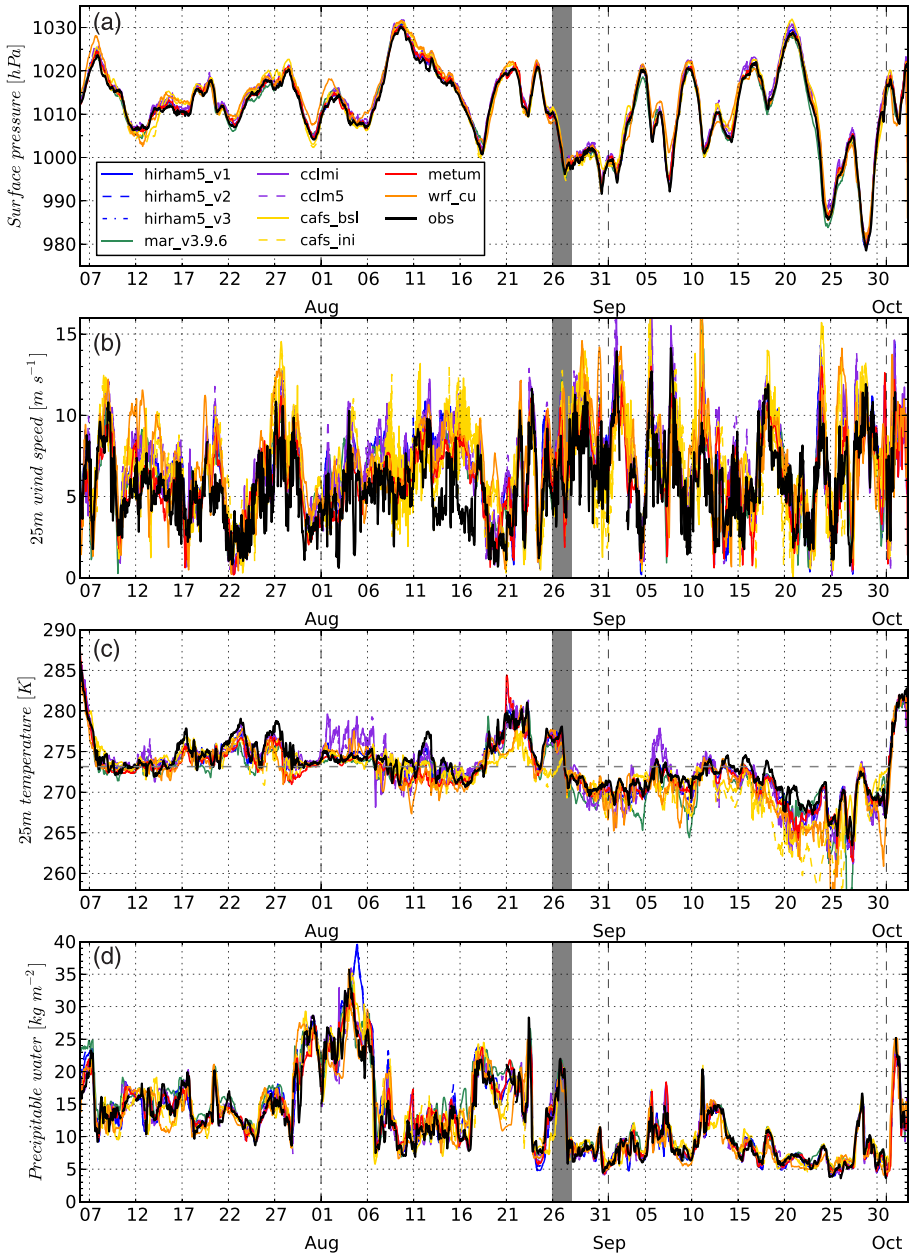

Figure 2. One-hour temporal evolution for the modeled (colors) and observed (black) (a) surface pressure [hPa], (b) wind speed [m/s] at 25 m above ground level, (c) temperature [K] at 25 m above ground level, and (d) column-integrated precipitable water [kg/m^2]. The gray-shaded vertical bar marks the time ±1 day of 27 August 2014, which Sotiropoulou et al. (2016) determined to be the transition between local melt season (prior) and freeze-up onset (after). The horizontal, dashed gray line in (c) marks the freezing temperature. The vertical, dashed black lines mark the first of each month. Note that the first month is July (not labeled).

Finally, near-surface turbulent SHF and LHF were estimated using eddy covariance. A 10-m mast was installed on a platform at the bow of the ship, reaching ~20 m above the surface. A heated Metek sonic anemometer measuring three-dimensional winds at 10 Hz was installed at the top of the mast, together with a Licor open-path gas analyzer, providing fast humidity measurements, and motion sensors. Computational fluid dynamics modeling of the flow distortion around the ship, together with careful measurements of ship's motions and accelerometer sensors collocated with the instruments, were used to correct and quality control the wind measurements and turbulent flux calculations (e.g., Moat et al., 2015; Prytherch et al., 2015).

3. Results

3.1. Near-Surface Meteorological Conditions

All comparisons are done to the observations, not to the forcing data. The forcing is only applied at the lateral boundaries and the free troposphere, not to the processes themselves. The forcing data itself—the reanalysis—are run in a data assimilation cycle, where short forecasts are corrected six-hourly using all available data, including some of the observations from ACSE. Hence, it would be unreasonable to expect any of the RCMs to outperform the reanalysis results, especially in the PBL. For that to happen, the RCMs would also have to be run with data assimilation, which would to some degree defeat the purpose of this study: to assess how well the RCMs can replicate the PBL structure, clouds, and energy fluxes important for the SEB given that the larger-scale development is reasonably accurate.

We therefore begin by examining variables that illustrate the meteorological evolution during ACSE. Along the ACSE track, a number of synoptic weather events were encountered, manifested in the surface pressure variability (Figure 2a). Large surface pressure variations were common, with increasing frequency but decreasing synoptic-event duration in late August through September, compared to July through mid-August. Low-pressure minima did not drop below 1,000 hPa until 27 August. Interestingly, Sotiropoulou et al. (2016) determined this date as the transition from summer melt toward autumn freeze season.

Near-surface wind speeds reflect the amount of synoptic variation observed during the ACSE transect. Wind speeds showed a variable range, from rather calm to over 10 m/s (Figure 2b). Often, the variability in wind speed followed the temporal variability in surface pressure. Variation in near-surface wind speed would be expected from different surface roughness lengths and elements associated with varying open water and sea-ice conditions, but the synoptic setting exerted a dominance on near-surface wind speed.

Seasonal variability is evident in the variations in 25-m temperature (Figure 2c). Near-surface temperatures before 27 August mostly remained at, or above, the freezing level, while after this, temperatures cooled substantially but remained between the freezing points of fresh and saline water until after around 20 September. PW decreased during cold higher-pressure events, while often showing localized peaks with low-pressure events (Figure 2d). PW also showed differences before and after 27 August; observed median PW decreased nearly twofold, from ~14 kg/m^2 before 27 August to ~8 kg/m^2 after. While the models did not always fully capture the 1-hr instantaneous observations (see Figures S1–S3 in the supporting information showing scatterplots of 25-m wind speed, temperature, and PW for each model), it is encouraging that all models reproduced the high- and low-frequency variability associated with the synoptic settings, suggesting that the general circulation is sufficiently constrained in the RCMs, despite the large integration domain used by the models. Furthermore, all models well captured the transitioning lower-troposphere thermodynamic state occurring around 27 August (gray-shade bar, Figure 2). Time-pressure cross sections of temperature and specific humidity from radiosoundings and models (Figures S4 and S5) corroborate the capacity of these models in reproducing the thermodynamic evolution of the troposphere observed during ACSE. Importantly, the models simulated the main weather events observed during ACSE, such as the warm/moist-air advection at the beginning of August and the seasonal thermodynamic transition in late August.

However, while the models capture the synoptic variability and transitions, vertical atmospheric structure exposes systematic deviations from the observations, summarized in Figure 3. The largest temperature and specific humidity errors are found across the midtroposhere to lower troposphere (Figures 3b and 3d). The temperature root-mean-square error (RMSE) is largest across the lower troposphere, decreasing with pressure. The specific humidity RMSE peaks at 850–925 hPa, with substantially lower values below and also tapering off with altitude. Many models show a cool bias near the surface up to ~900 hPa, with smaller systematic errors through the free troposphere (Figure 3a). Most of the models had a dry bias in specific humidity at the lowest level (Figure 3c), where temperatures were too cool. In general, cafs and wrf_cu (the atmospheric model in both RCMs is WRF) were the coldest and driest in the lowest troposphere, while mar_v3.9.6 had the largest moist bias, persisting across most of the troposphere. The other models generally exhibited smaller mean bias error and RMSE profiles in both temperature and humidity.

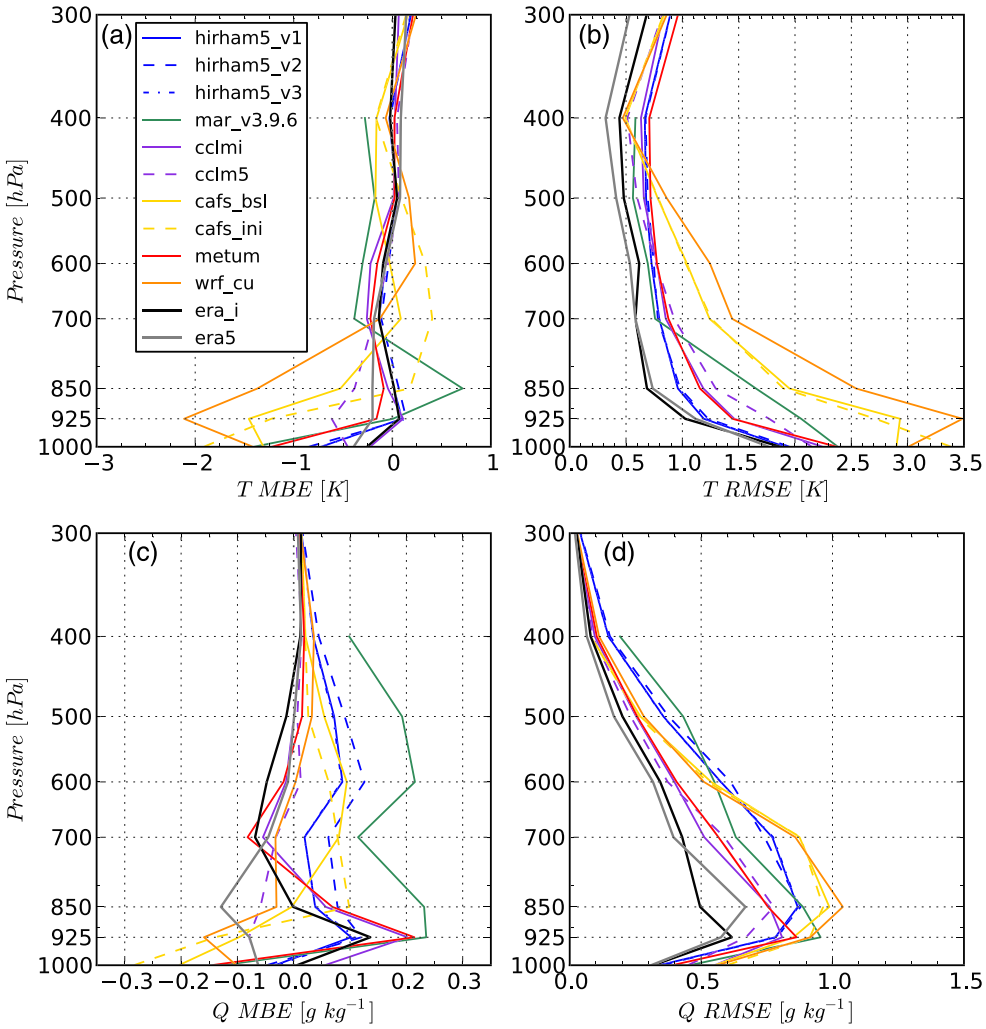

Figure 3. (a and b) Vertical temperature ([K]) and (c and d) specific humidity ([g/kg]) profile statistics, where mean bias error (MBE) are shown in panels a and c and root-mean-square error (RMSE) in panels b and d. Models and observations were linearly interpolated onto standard pressure levels of 1,000, 925, 850, 700, 600, 500, 400, and 300 hPa. Reanalysis error profiles are included for ERA-Interim (black) and ERA5 (gray). Statistics are valid for the full Arctic Clouds in Summer Experiment (ACSE) campaign, 5 July at 01:00 UTC through 3 October at 07:00 UTC.

For comparison, the bias statistics from ERA-Interim and ERA5 reanalyses relative to the observed profiles are also shown in Figure 3 (black and gray lines). Reanalysis RMSEs for temperature and specific humidity were substantially lower than for all RCMs across the whole troposphere. This is expected as the reanalysis assimilated all the ACSE radiosounding measurements, in addition to other near-surface meteorological observations onboard the icebreaker. The reanalyses may also have benefitted from assimilation of other surface observations in the region as well as of satellite observations from the vicinity of the icebreaker. This information is unavailable to the RCMs except through the nudging.

For both temperature and specific humidity, mean bias error profiles for the RCMs were similar in sign and shape to the bias profiles of the reanalysis product that were used to nudge or initialize the forecasts (Figures 3a and 3c). This result indicates that errors in the background thermodynamic state from the reanalyses may be emergent within the domain of the RCMs; exceptions include the mar_v3.9.6, cafs, and wrf_cu simulations, revealing much larger biases across the lower atmosphere. However, the fact that that the RMSE is systematically larger in all RCMs indicates that the effect of the nudging does not inhibit the RCMs to develop their own thermodynamic state. Without the forcing from reanalysis, the RCMs would not be able to simulate the true atmospheric evolution with relatively high accuracy (Figures 2 and S1–S3), which would render an intercomparison to observations useless.

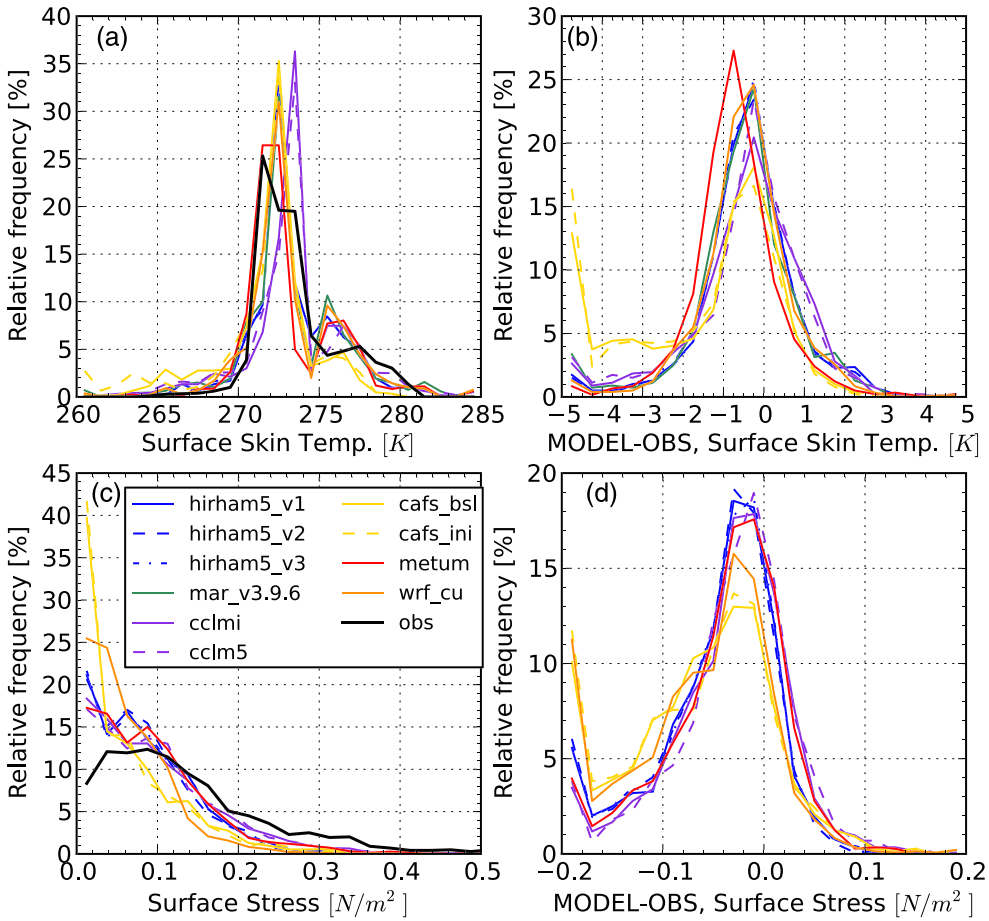

Figure 4. Relative frequency distributions of the (a) surface skin temperature [K], (b) surface skin temperature model minus observation difference [K], (c) surface stress [N/m^2], and (d) surface stress model minus observation difference [N/m^2]. Note that mar_v3.9.6 was unable to extract surface stress. Statistics are valid for the full Arctic Clouds in Summer Experiment (ACSE) campaign, 5 July at 01:00 UTC through 3 October at 07:00 UTC.

It is important to note that the biases vary temporally, illustrated in time-pressure cross-section differences (Figures S6 and S7). All models shared a common cold bias in the lower troposphere (Figure S6). The bias was exacerbated between late July and mid-August in connection with a persistent pulse of warm air with above-freezing temperatures (Tjernström et al., 2015; Figure S4). The low-level cold bias was generally eliminated after this event and through the remainder of the simulation, except for in wrf_cu and cafs. Specific humidity errors were also largest prior to 27 August (Figure S7), especially in connection with the advection event. Splitting the simulations to periods before and after 27 August, RMSEs in specific humidity during the latter were a factor of 2 smaller than those in Figure 3d (not shown).

The observed surface skin temperature relative frequency distribution (RFD) is bimodal, with a broad dominating peak across the freezing points of fresh and saline water, around 271 to 273 K, and a weaker peak a few degrees above freezing (Figure 4a). Modeled skin temperatures show a similar bimodality but with narrower distribution peaks. Difference distributions indicate that the models were most often too cool by 0.5 to 1 K (Figure 4b); the cool bias was slightly larger for metum. The various methods of prescribing sea-ice concentration among the models (Table 1) yielded intermodel differences from the observed AMSR2 daily satellite-observed sea-ice concentrations (Figure S8). The modeled sea-ice concentrations (SICs) in cafs, and to a lesser extent in wrf_cu, were frequently larger than 80%, and as such, caused the relatively high frequency of skin temperature differences below −3 K (Figure 4b). The overly high SICs in these models may also have contributed to the lower tropospheric cold biases found in Figures 3 and S6f and S6i–S6j.

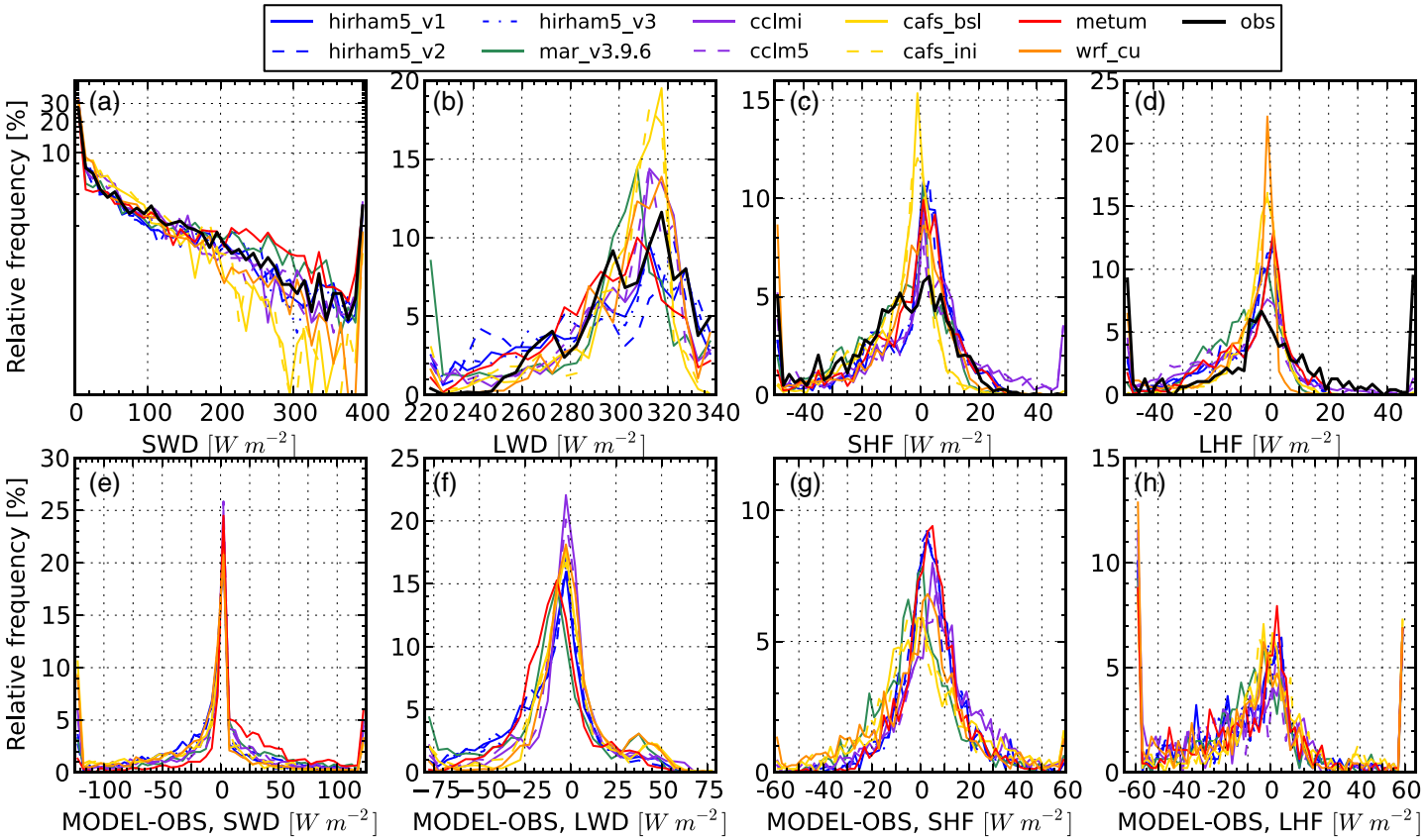

Figure 5. Observed (black) and regional climate model (colors; see legend) relative frequency distributions of (a) shortwave downwelling (SWD; W/m^2), (b) longwave downwelling (LWD; W/m^2), (c) sensible heat flux (SHF; W/m^2), (d) latent heat flux (LHF; W/m^2], and model minus observation relative frequency distributions of (e) SWD (W/m^2), (f) LWD (W/m^2), (g) SHF (W/m^2), and (h) LHF (W/m^2). Statistics are valid for the full Arctic Clouds in Summer Experiment (ACSE) campaign, 5 July at 01:00 UTC through 3 October at 07:00 UTC.

In terms of surface stress, the observations reveal a broad peak distribution between 0.04 and 0.1 N/m^2, including a relatively long, positive tail (Figure 4c). All models underestimated surface stress and the observed distribution shape by having the most frequently occurring stress at the smallest values, followed by a sharper distribution drop off for increasing stress; exceptions were the hirham5 and metum simulations, that did manage to recover a secondary peak near that of the observations. The generally large underestimation in modeled surface stress for wrf_cu and cafs is consistent with an overrepresentation of ice-covered surfaces, which in principle, would have a smaller roughness length compared to open water. It is likely the combination of too much sea ice and too low surface stress contributed to overestimating the near-surface winds in all models, and especially in cafs and wrf_cu (Figure S1). As cafs was the only fully coupled atmosphere-sea-ice model, the exaggerated production of sea ice, and hence colder surface temperatures and reduced wind stress, is likely connected to errors in the SEB; these connections and associated biases will be explored in the next section.

3.2. Surface Energy Fluxes

In a broad sense, RFDs of the models' SEB components bear similarity to those observed during ACSE (Figure 5). Relatively small SWDs (<100 W/m^2) were most frequent, LWD typically ranged 290–330 W/m^2, and turbulent heat fluxes most frequently between ±40 W/m^2; SEB fluxes are defined positive when energy is transferred from the atmosphere to the surface.

While the models captured these dominant, observed ranges, subtle statistical discrepancies in the radiative fluxes indicate model-specific biases. For example, both cafs simulations and, to a lesser extent wrf_cu, underestimated large SWD values, while metum and mar_v3.9.6 overestimated SWD occurrences above

200 W/m^2 (Figures 5a and 5e). Peak LWD distributions were narrower for cafs and mar_v3.9.6 than for the other RCMs, also containing sharper declines in frequency above 320 and 310 W/m^2, respectively. All but the hirham5 model simulations underestimated the occurrence of LWD greater than 320 W/m^2. However, all models overestimated LWD less than 250 W/m^2. Because of overestimated small, and underestimated large, LWD values, the error RFDs peaked below 0 W/m^2 and were negatively skewed (Figure 5f); the distributions for metum and mar_v3.9.6 stand out, shifted further negative; the latter substantially overestimated LWD below 230 W/m^2 (Figure 5b). Downwelling radiative fluxes are strongly affected by the presence and microphysical properties of clouds. These results suggest that some models that systematically (metum and mar_v3.9.6) overestimate SWD and underestimate LWD simulate too few clouds, or too little liquid condensate in clouds, while other models underestimating SWD and overestimating LWD (cafs and wrf_cu) likely overrepresent liquid-bearing cloud cover (e.g., Wyser et al., 2008). Furthermore, while the distributions for hirham5 and cclm simulations generally agree with the observations (Figures 5a and 5b), the error distributions indicate that instantaneous errors in SWD and LWD are still present (Figures 5e and 5f).

Net turbulent heat fluxes over the summertime Arctic Ocean are typically small (Persson et al., 2002; Sedlar et al., 2011; Tjernström et al., 2005). However, the distribution shape and magnitude of modeled errors here (Figures 5c and 5d and 5g and 5h) were similar in size to the actual heat fluxes (e.g., Tjernström et al., 2005). SHF error distributions generally peaked around 0 W/m^2; hence, modeled fluxes have only a small, systematic statistical error. However, the shape of the SHF-error RFD is quite similar to the RFD for the fluxes themselves, so that random errors are similar in magnitude to the flux itself. Generally, the modeled SHF errors revealed a negative correlation to observed SHF (Figure S9). This result implies that random errors in the modeled surface and near-surface temperatures (Figure 4b), together with deviations in wind speed (Figure S1), contributed to errors in SHF (Tjernström et al., 2005). The shapes of the error distributions can also be separated into groups. The cafs, wrf_cu, and mar_v3.9.6 distributions were slightly shifted and skewed toward negative values (energy from surface to atmosphere); metum and hirham5 simulations had narrow difference distributions, while the cclm distributions were slightly shifted and skewed toward positive values (Figure 5g). Small but negative SHFs dominated the cafs simulations all, while positive SHFs were underestimated (Figure 5c); wrf_cu also frequently overestimated larger negative fluxes. These are the same models that showed the largest near-surface and lower troposphere cold bias (Figure 3), implying a relationship.

Measurements of LHF in polar regions are notoriously difficult (e.g., Persson et al., 2002), and here proved the same, with an observed LHF recovery rate of only 21% recovery (466 of 2,168 hr) for 1-hr averages. Consistent with the models' SHF distributions, their LHFs were also frequently too largely negative (Figures 5d and 5h), resulting in an overestimation of moisture transport from the surface to the air.

To explore causes for differences in SHFs, we turn to the bulk parameterization. $\mathrm{SHF_{bulk}}$ combines the wind speed at a reference level (U_{ref}) with the gradient between the surface and near-surface air temperatures (ΔT), multiplied by a heat exchange coefficient (C_H), itself a function of stratification, to estimate SHF:

$$\mathrm{SHF_{bulk}} = \rho_{\mathrm{air}} c_p U_{\mathrm{ref}} C_H \Delta T, \tag{1}$$

and includes the density of air (ρ_{air}) and specific heat capacity of dry air (c_p). We can rearrange the terms in this parameterization such that SHF is scaled by U_{ref} and plot this against the near-surface ΔT, using surface and reference level (10–12 m AGL depending on the model's height of each model's lowest level) values in equation (1) from the respective RCMs and observations (Figure 6).). The slope of the distribution is then proportional to C_H and therefore moderates the transfer of sensible heat in the surface flux parameterization scheme. Scaled SHFs from the observations were largest when the surface was warmer than the near-surface air temperature and decreased following a linear dependence with decreasing ΔT (black circles). The broadly near-linear slope for $\Delta T < 0$ K distributes around the linear regression with a positive slope (gray line), indicating that C_H was relatively constant in this regime. The distributions for all RCMs besides cafs (panels g and h) managed to resolve the positive, linear dependence of scaled SHF for $\Delta T < 0$ K, indicating that errors in modeled SHFs for this temperature difference regime are likely not stemming from the surface flux parameterization scheme. Instead, what we do find is that the scaled fluxes for the models are more uniform and generally lack the variability that was observed, pointing toward differences in the near-surface winds (e.g., Figure S1) and errors in surface (e.g., Figures 4a and 4b) and near-surface temperatures

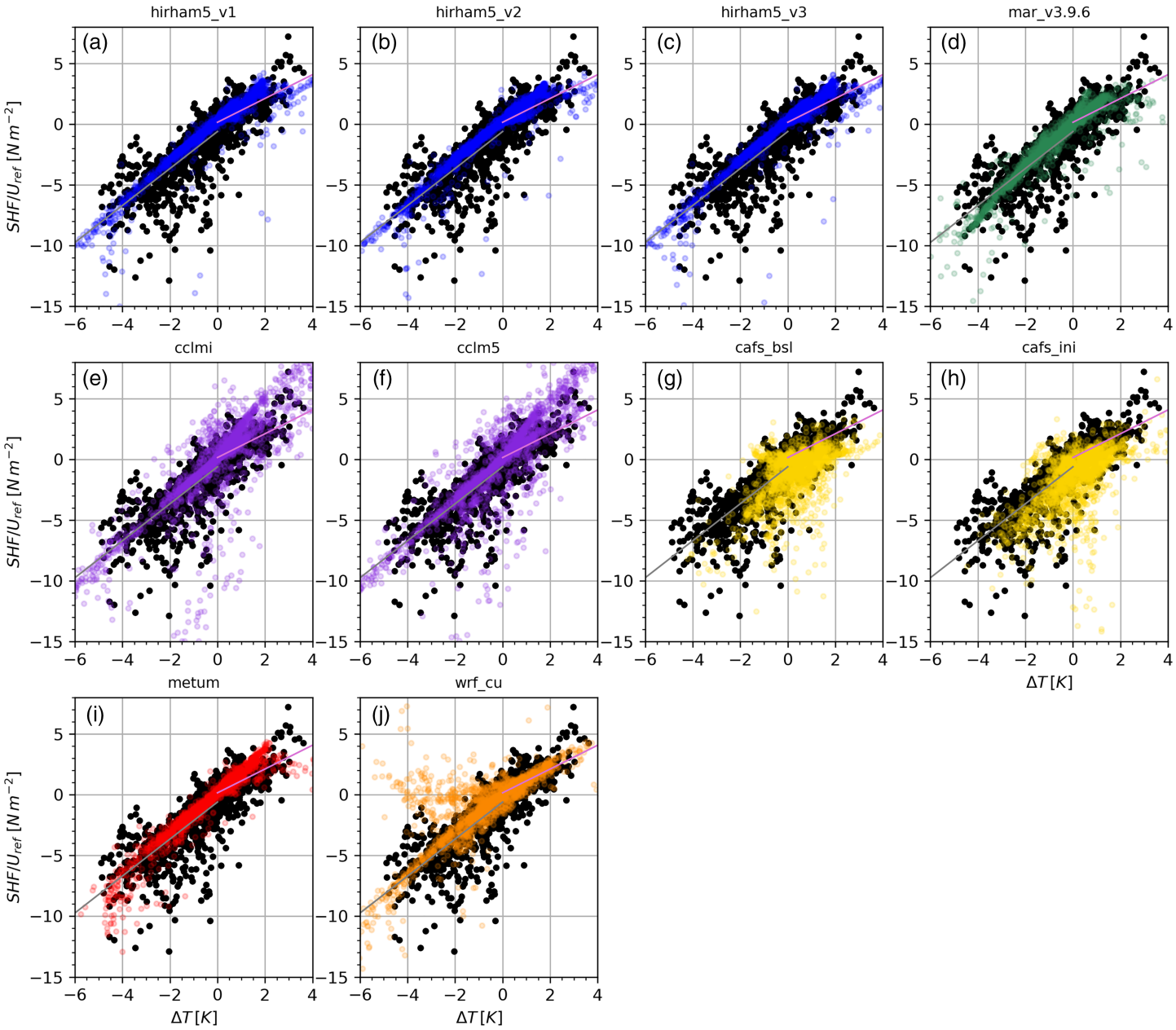

Figure 6. Covariability between the sensible heat flux (SHF, W/m^2) weighted by the reference-level (10–12 m, depending on model) scalar wind speed (U_{ref}, m/s) and the near-surface (10–12 m, depending on model) air temperature minus surface temperature difference (ΔT, K) for observations (black) and each regional climate model (colors, see title for model name). The gray line is the linear regression for the observations of scaled SHF for $\Delta T < 0$ K; the pink line is the linear regression for observations when $\Delta T > 0$ K. Statistics are valid for the full Arctic Clouds in Summer Experiment (ACSE) campaign, 5 July at 01:00 UTC through 3 October at 07:00 UTC.

(Figure S6) as the cause; Tjernström et al. (2005) reported a consistent result from SHEBA simulations. The cafs simulations better simulated variability in the scaled SHF; however, these distributions do not resolve a clear, positive relationship with ΔT, especially for cafs_bsl (panel g).

When $\Delta T > 0$ K, the observations reveal a slightly different regime where the scaled SHF does not increase as dramatically with ΔT and so the slope, while still positive, is reduced (pink line). Tjernström et al. (2005) noted a change in slope for the positive temperature difference regime, and all but one RCM in their study failed to reproduce it. Here we find relatively modest to moderate changes in the models' distribution of scaled SHF when ΔT was stably stratified. The more modest slope changes were simulated by hirham5,

mar_v.3.9.6, and metum. Besides mar_v3.9.6, these same models consistently show a peak occurring at more positive SHFs compared to the observations (Figure 5g), suggesting the lack of slope decrease has contributed to a slightly larger and more efficient heat exchange under the stable ΔT regime. CCLM simulations reveal a split in the scaled flux distributions, where the slope for one branch follows a similar slope as found when $\Delta T < 0$ K, while another branch shows a more amplified, positive slope (panels e and f). Consequently, the scaled fluxes under near-surface stable stratification in cclm are connected to overestimation of the exchange coefficient. As a result, the SHF distributions for cclm were skewed toward larger, positive values contributing to relatively large additional +10 to +40 W/m^2 energy input into the surface (Figures 5c and 5g). Further analysis revealed the amplified-positive slope branch in Figures 6e and 6f occurred during hours when the sea-ice concentration was above 90% (not shown), indicating that the surface roughness parameterization may be contributing to the overly excessive exchange of sensible heat in this model. With cafs, we find the positive SHFs for $\Delta T > 0$ to be relatively small, and instead, its distributions are dominated by negative SHFs (panels g and h), which agrees with its frequency distributions in Figures 5c and 5g. Such a counter-gradient aspect to the SHFs for cafs suggests that errors are emerging between mean temperature structure and the treatment of flux contributions from open leads or melt ponds within a grid cell.

3.3. Factors Impacting Surface Energy Fluxes

3.3.1. Cloud Water and Its Vertical Distribution

RFDs for cloud LWP (Figures 7a and 7b) and IWP (Figures 7c and 7d), and model-observed scatterplots (Figures S10 and S11), provide some insight into the modeled surface radiative fluxes. Observations show decreasing frequency of occurrence with increasing LWP and IWP, a behavior the RCMs resolved. However, the degree to which models mimicked the decreasing frequency of higher LWPs varied greatly, in both an intermodel sense and a temporal sense. The cafs and wrf_cu simulations largely overrepresented occurrences with LWP > 100 g/m^2, primarily at the expense of the more frequently observed 0–100 g/m^2 range; the occurrences of LWP 350 g/m^2 for these models was nearly as large, or larger, as their cases of LWP < 50 g/m^2. In contrast, LWPs in mar_v3.9.6 and metum were nearly exclusively below 100 g/m^2, especially during the colder and drier second period (Figures S1 and S2). During the latter period, hirham5 and cclm simulations also deviated from the observed distribution, underestimating clouds with LWP > 25 g/m^2 (Figure 7b). However, the cafs_ini, with increased IN concentrations, dramatically reduced the occurrence of substantial cloud liquid in the second period, in better agreement with observations than cafs_bsl. LWP scatterplots and correlation coefficients indicate that while the distribution shapes modestly agreed, instantaneous values between models and observations were relatively poor (Figure S10), a key result also found for the ARCMIP RCMs simulating the SHEBA experiment (Wyser et al., 2008).

During both periods, observed IWP was most frequently below 50 g/m^2 (Figures 7c and 7d). RCM IWP distributions reflected the observed distribution shape, although the majority of the models overestimated cloud ice for both periods, especially for values above 20 g/m^2; overestimated frequencies of IWP came at the expense of underestimating IWP between 0 and 20 g/m^2 (Figure S11), especially during the latter period (Figure 7d). The hirham5 and mar_v3.9.6 simulations consistently had the highest frequencies of the largest IWPs. Oppositely, the cclm simulations did not produce IWPs above 100 g/m^2. Interestingly, the cafs_ini IWP distributions were not very different from the cafs_bsl, although the increased IN concentration did cause a reduction in the cloud LWP during the latter period (Figure 7b). This suggests that the conversion of cloud liquid to ice, and subsequent ice crystal growth, resulted in increased frozen precipitation from the atmosphere to surface, and hence, the increased ice did not remain within the cloud.

Shortwave transmissivity (ST) is calculated as the ratio between the surface SWD and the SWD at the top of the atmosphere, which was estimated for each 1-hr profile using a radiative transfer model (Mlawer et al., 1997). ST provides a measure of how much shortwave radiation was scattered, reflected, or absorbed within the atmospheric column, with small ST indicating large atmospheric attenuation of SWD. The differences in the modeled cloud water paths and SWD (Figures 5a and 5b) RFDs are further reflected in the ST distributions. Especially for the period 5 July to 25 August, when solar zenith angles were smaller than later, models that overestimate LWP (cafs & wrf_cu) also underestimate the transmissivity (Figure 7e); ST was considerably larger for the models that underpredicted LWP (mar_v3.9.6 and metum). The connection between LWP and IWP and ST suggests discrepancies between modeled and observed cloud properties. Interestingly, the

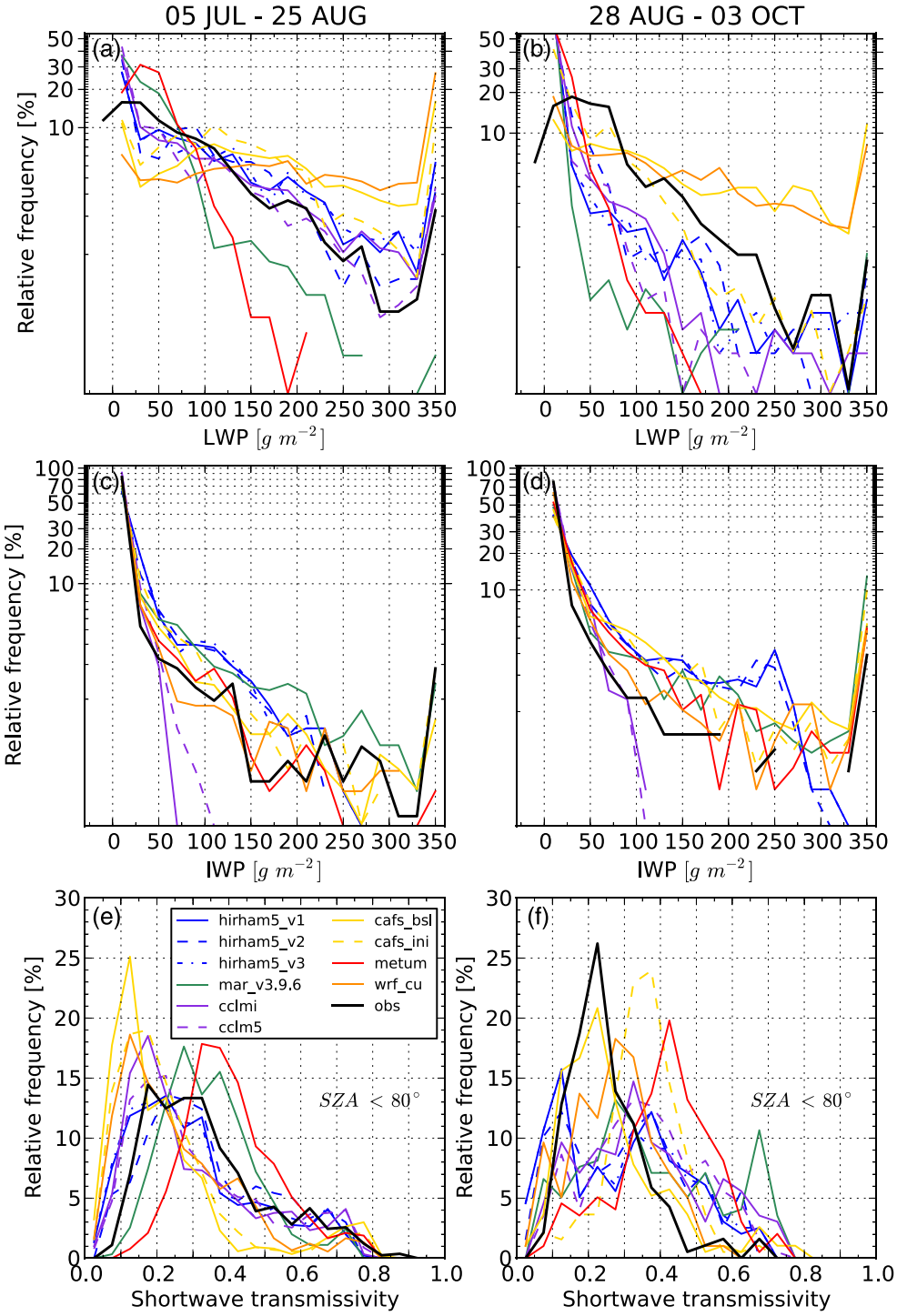

Figure 7. Relative frequency distributions of (a and b) liquid water path (LWP; g/m^2), (c and d) ice water path (IWP; g/m^2), and (e and f) shortwave transmissivity (ST) for the models and observations. The distributions in panels a, c, and e are valid for 5 July through 25 August, while panels b, d, and f are for 28 August through 3 October. ST is computed as the following ratio: SWD_{sfc}/SWD_{toa}, and only for solar zenith angle (SZA) smaller than 80°; the SZA requirement limited the ST distributions during the second period to only 22% of the full period.

change in IN concentration between the cafs runs had a profound impact on ST during the second period (Figure 7f) but had a minor impact during the first. Sotiropoulou et al. (2016) reported more warm-phase clouds during the first period, compared to the second, agreeing with >80% frequency distribution of

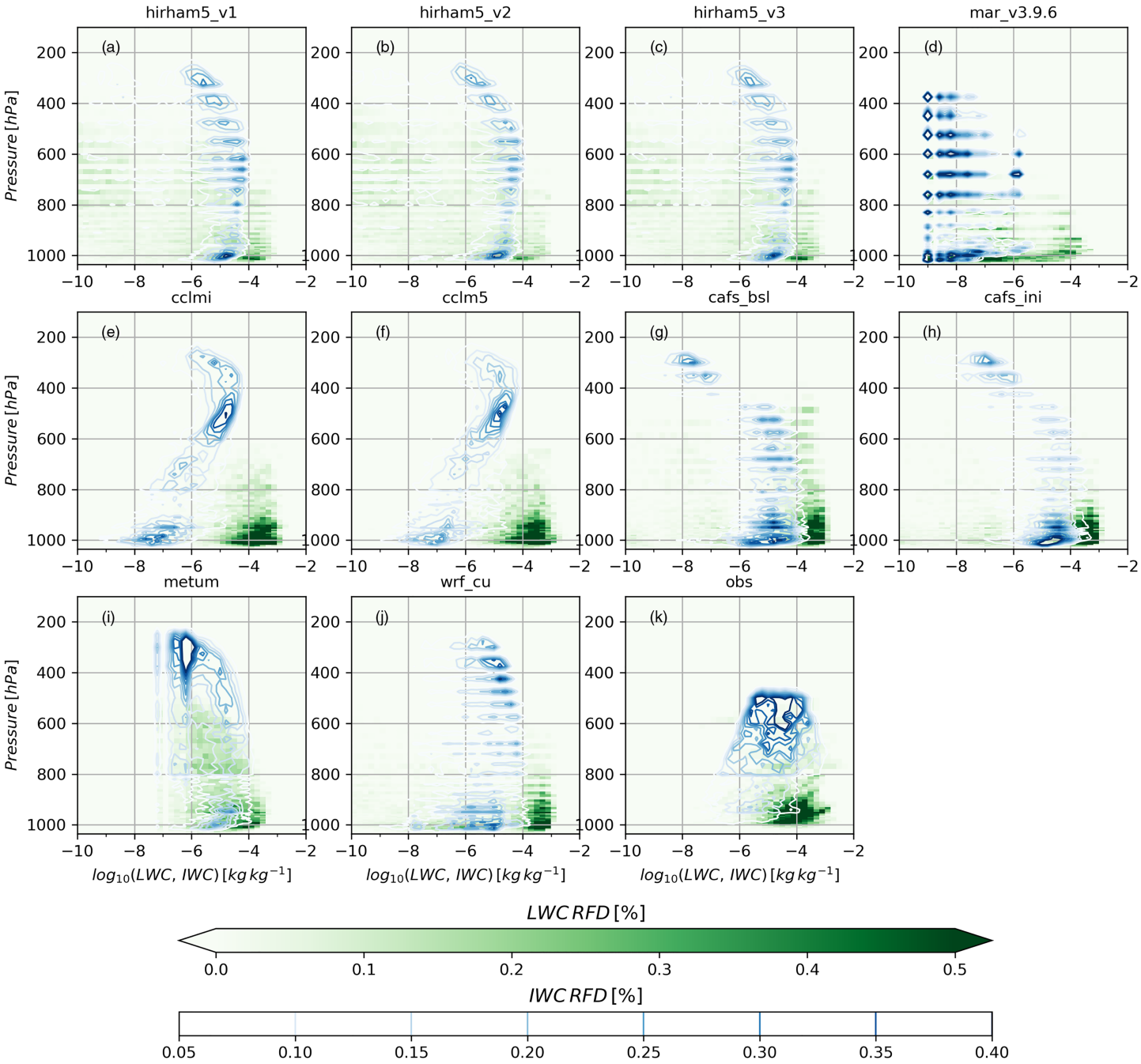

Figure 8. Relative frequency distributions of vertical profiles of cloud liquid water content (LWC, kg/kg, green contour shading) and ice water content (IWC, kg/kg, blue contour lines). Regional climate models are shown in panels a–j, with the model name in the title, while observations are in panel k. The values of LWC and IWC have been taken to the log10 (*x* axis). Note that the observed retrievals of LWC and IWC were limited to the highest vertical resolution of the cloud radar, 6,000 m AGL, which is corresponding to approximately the 500-hPa pressure level. Statistics are valid for the full Arctic Clouds in Summer Experiment (ACSE) campaign, 5 July at 01:00 UTC through 3 October at 07:00 UTC.

IWPs during the first period (Figure 7c); the latter period was dominated more by the prototypical Arctic mixed-phase clouds. The change in ST distributions for cafs simulations is consistent with increasing solid-phase cloud hydrometeors in cafs_ini during the second period, resulting in a larger consumption of cloud liquid into ice. But since IWPs for cafs_ini were similar in distribution to cafs_bsl (Figures 7c and 7d and S11g and S11h), the change in ST is mainly due a lower cloud optical depth in connection with reduced cloud liquid.

To investigate possible causes for the statistical grouping of models by their water paths, we examine the vertical distribution of LWC and IWC in Figure 8. Focusing first on IWC (blue contour lines), we find unique distribution profiles among the different models. There is a wide range in predominant IWCs across the lower troposphere, spanning multiple orders of magnitude among the models. The mar_v3.9.6, cclm, and wrf runs (panels d–f and j) have the least amount of IWC in the lower atmosphere, with the other remaining models having more similar magnitudes. Retrievals of IWC from the radar suggest that the models with lower IWCs in the lower troposphere are more realistic. Figures S12 and S13 separate the vertical hydrometeor distributions into the two periods, which reveal that more ice was observed in the lower troposphere during the colder, second period, a feature the models captured. The profile shape of IWC across the free troposphere also varied. Observations reveal a broad range of IWCs spanning about 2 orders of magnitude across the most of the free troposphere (panel k); the radar had a maximum vertical extent of 6,000 m and so no retrievals exist above about 500 hPa. All models besides cclm had a consistent, broad distribution range of IWC across the troposphere to the 500-hPa level, although the IWC ranges spanned one (panels a–c) to over 3 orders of magnitude (panels d, i, and j). The profile of IWC in cclm generally increased from very low ice near the surface to its largest IWCs in the upper free troposphere (panels e and f).

Cloud liquid (green contour shading) was most frequent at pressures larger than 800 hPa (Figure 8), and all models simulated the preference for low-level liquid-bearing clouds. However, the magnitude of LWCs varied by orders of magnitude among models and with observations. The cclm simulations were most similar to the observations in representing the spread and vertical partitioning of LWC (panels e and f). For hirham5, cloud liquid was present across most of the troposphere, although the LWCs were often very small (panels a–c). Simulations of LWP from both of these models were most similar to the observed distribution during the first period (Figure 7a). Considering their different vertical partitioning of LWC, cclm and hirham5 simulated rather different clouds at any given time. Additionally, only modest differences between the different microphysical parameterizations employed in the hirham5 simulations are present. The mar_v3.9.6 (panel d) and metum (panel i) models that underestimated LWP (Figures 7a and 7b) do so because of an underestimation of low-level liquid. For metum, the underrepresentation of cloud liquid is because too much ice is formed at the expense of supercooled liquid (Furtado & Field, 2017). Oppositely, cafs and wrf_cu show large and relatively consistent LWCs distributed vertically below 800 hPa (panels g and h and j). When IN concentrations were increased (panel h), the vertical extent of cloud liquid decreased relative to the baseline simulation (panel g), but the lower level LWCs largely remained unchanged between cafs runs. Therefore, the decrease in LWPs between the cafs_bsl and cafs_ini (Figures 7a and 7b) appears to be connected to a reduction in liquid-bearing clouds in the free troposphere.

3.3.2. Cloud LWP-Radiation Interaction

Here we investigate how cloud longwave emission temperature varies with cloud emissivity for low liquid-bearing clouds, with a focus on the layer between the surface and 800 hPa, where cloud liquid was more frequently found in both observations and models (Figure 8). Observed retrievals of LWC were integrated from the first radar range (~70 m AGL) to the height level corresponding to the 800-hPa pressure level, providing the LWP of these low cloud layers. We then sampled only instances where liquid hydrometeors were present across this layer and the total LWP across this layer was at least 5 g/m^2 and then apply the bulk parameterization to estimate the low cloud layer emissivity, ε

$$\varepsilon = 1 - e^{(-k_c \cdot \mathrm{LWP})} \tag{2}$$

with k_c as the mass absorption coefficient value of 0.158 m^2/g (Stephens, 1978). The observations (black circles) in Figure 9 depict the range of low cloud layer emissivity when at least 5 g/m^2 of integrated liquid were present, showing a dominant mode of black or near blackbody ($\varepsilon \approx 1$) emitting clouds. The blackbody (or black) clouds are observed to occur across a range of effective infrared emission temperatures, defined as the atmospheric temperature where the integrated LWC from near surface up to 800 hPa reaches a value of 5 g/m^2; this temperature can be approximated as the liquid cloud base temperature. Graybody (or gray) clouds ($\varepsilon < 1$) were also frequent and across the same range of emission temperatures as the black clouds (Wyser et al., 2008), indicating the rather frequent occurrence of supercooled liquid.

Emission temperature and effective emissivity for the modeled low-level clouds were computed as in the observations from the profiles of LWC and temperature. Here we subsample the models to the times

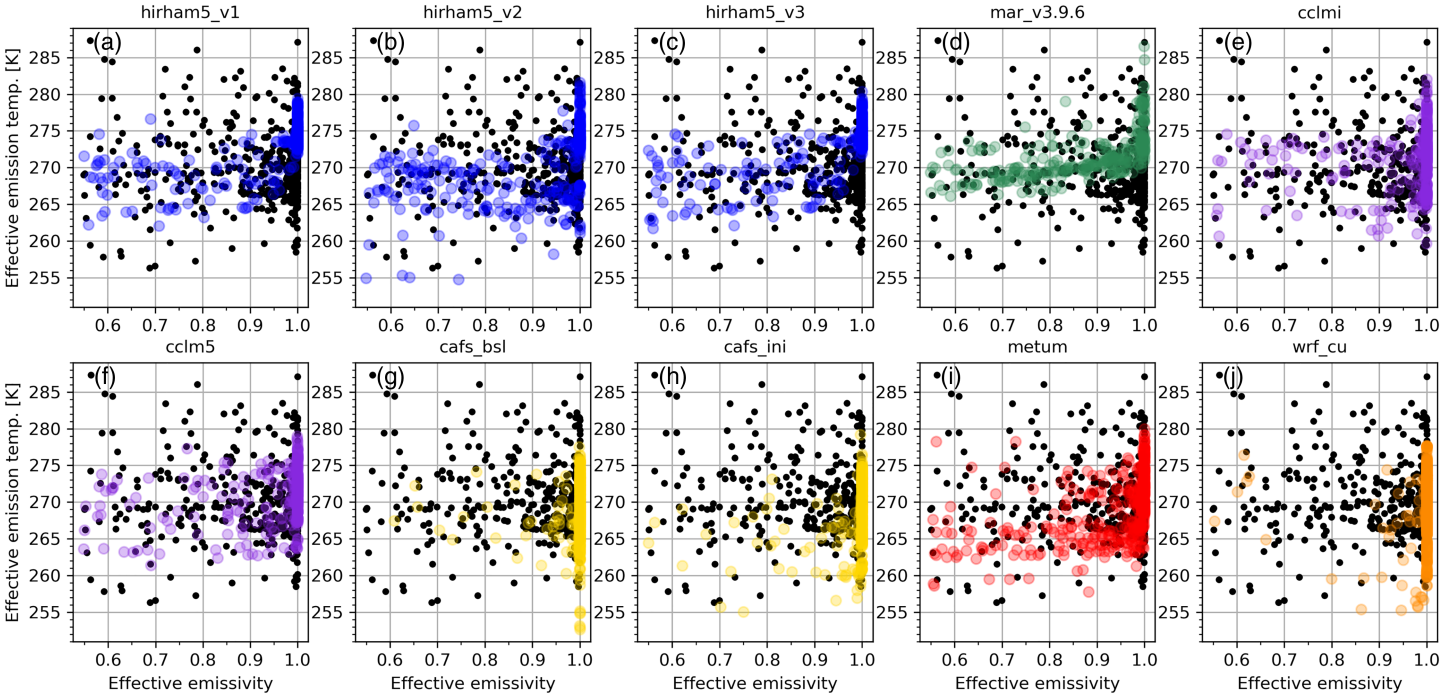

Figure 9. Covariability between low-level (pressure > 800 hPa) liquid cloud effective emission temperature (K) and the longwave effective emissivity estimated from equation (2), for observed low clouds with at least 5 g/m^2 of liquid water path. Observations are in black, and the regional climate models are colors, with the name of each model in the panel title. See text for explanation of the derivation of low cloud liquid water path, effective emissivity, and cloud emission temperature.

when a low-level cloud emissivity and emission temperature were observed; the modeled distributions overlay the observed (colors in Figure 9). There are clear differences in the physics of the low-level liquid bearing clouds among the models. The wrf_cu (panel j) and cafs simulations (panels g and h) nearly exclusively represent low clouds as near blackbodies; very few gray clouds are simulated in these models, but if they are, they occur at temperatures below freezing. All other models reveal a combination of black and gray clouds, but with differing characteristics. The mar_v3.9.6 (panel d), hirham5_v1 (panel a), and hirham5_v3 (panel c) fail to simulate clouds with $\varepsilon \approx 1$ at temperatures below freezing, suggesting an artificial conversion of liquid to ice in these models. However, there are more supercooled liquid clouds with $\varepsilon \approx 1$ for hirham5_v2 (panel b), implying that the Tompkins microphysics parameterization scheme is better suited at partitioning liquid and ice for low clouds compared to the Sundqvist et al. parameterization in v1 and v3 simulations. This version also contains more liquid clouds at freezing temperatures with emissivities below 0.9 compared to versions with Sundqvist microphysics. The distributions for the cclm (panels e and f) and metum (panel i) runs capture the temperature range of blackbody clouds and also simulate graybody clouds across a range of emission temperatures. There are a number of cloudy conditions with emission temperatures around 260 K and for a range of emissivity values in metum. The colder emission temperatures spread across the lower emissivity range appear to be connected to the model's overly efficient production of ice at the expense of cloud liquid across the lower atmosphere (see Figure 8i). Finally, all models fail in simulating clouds with emission temperatures warmer than approximately 275 K and $\varepsilon < 1$. The models are able to represent these warmer emission temperatures; however, they are almost exclusively modeled with an $\varepsilon = 1$. The larger emissivity and warmer temperatures of these clouds in the model runs will contribute to overestimated LWD compared to observations for these cases.

The median and interquartile distributions of LWD and SWD for particular observed and modeled LWP ranges are shown in Figure 10. Cloud emissivity increases with LWP, resulting in an increase in LWD, until the cloud emissivity approaches a blackbody (e.g., Stephens, 1978), typically within a LWP range of 30–50 g/m^2 (Shupe & Intrieri, 2004). After becoming a blackbody, variations in LWD are connected to emission temperatures and cloud droplet size effects (Garrett et al., 2002; Garrett & Zhao, 2006; Lubin & Vogelmann,

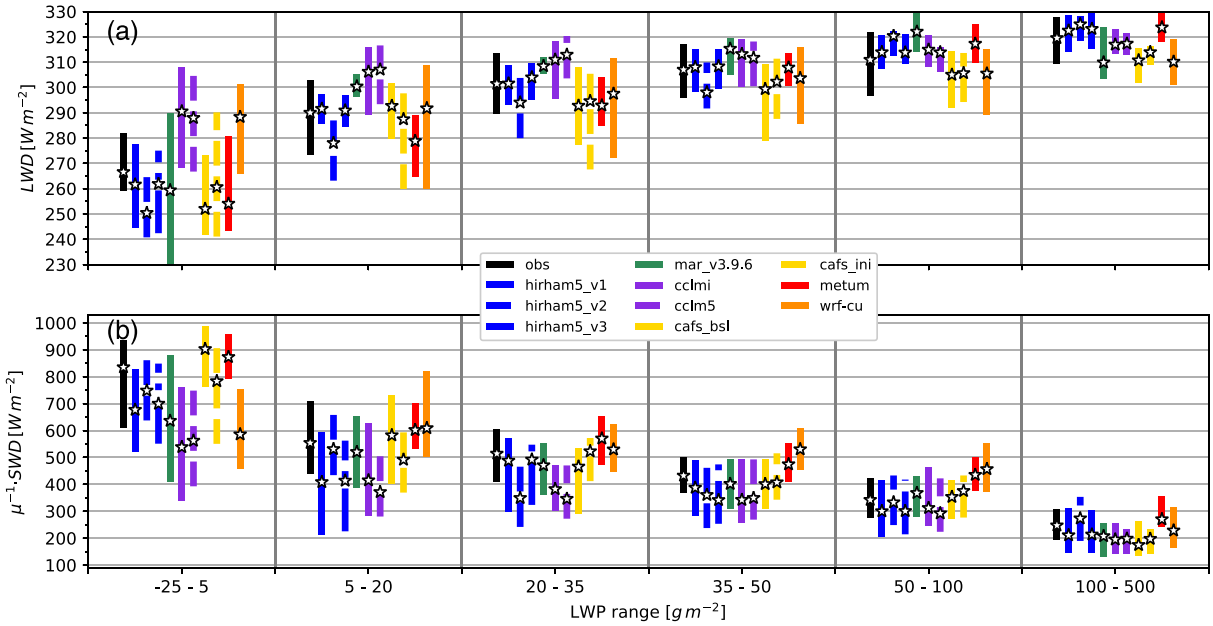

Figure 10. Interquartile ranges (bars) and median (a) longwave downwelling (LWD; W/m^2) and (b) (μ^{-1}) * shortwave downwelling (SWD; W/m^2] for six selected ranges of LWP [g/m^2]. SWD in (b) is weighted by the cosine of the solar zenith angle, μ; only periods when the solar zenith angle <80° are included. Statistics are valid for the full Arctic Clouds in Summer Experiment (ACSE) campaign, 5 July at 01:00 UTC through 3 October at 07:00 UTC.

2006). Conversely, the optical thickness of the cloud layer also increases with LWP, resulting in enhanced solar extinction and a decrease in surface SWD. Note that since each model's LWD/SWD is grouped using the model's own LWP, this metric isolates each models' cloud-radiative interaction at a given LWP, regardless of whether or not that instantaneous LWP was correctly simulated.

The RCMs managed to capture these broad, expected cloud-radiation interactions. However, intermodel differences are apparent. CCLM simulated relatively large LWD for low LWPs (< 35 g/m^2), while these clouds also amplified the SWD reduction. This behavior is difficult to understand considering the cloud emissivity and emission temperatures generally agreed reasonably with the observed distribution (Figures 9e and 9f). It is possible that the relative sparsity of low-level cloud ice in these simulations fails to interact significantly with the persistent cloud liquid (Figures 8e and 8f). As such, the radiative transfer through the cloud is only influenced by the cloud droplets with only small variability in the longwave introduced entirely through increases in LWP. In a similar, fashion, median, and quartile ranges of both LWD and SWD in wrf_cu were relatively invariant to the amount of LWP compared to the other models; obviously, the dominance of near blackbody clouds in the lower atmosphere is the reason (Figures 9j and 8j). The sensitivity in SWD to LWP in metum was also lower, although its LWD was more sensitive to the amount of cloud liquid. Cloud-radiative interactions in the hirham5_v2 simulation, with the Tompkins cloud scheme, differed from hirham5 simulations using the Sundqvist scheme (v1 and v3), with less LWD and more SWD for nonblackbody emitting clouds. This result is consistent with the results in Figures 9a–9c where hirham5_v2 exhibited the ability to maintain lower cloud emissivity at slightly cooler emission temperatures. The presence of less opaque and cooler clouds should reduce the amount of LWD while also decreasing the extinction of SWD for low LWP clouds; changes in cloud cover and phase partitioning, including changes to atmospheric temperature structure, have been attributed to the choice of cloud scheme in hirham5 (Klaus et al., 2016). For cafs, the differences in radiation with LWP are difficult to explain. At LWPs above 35 g/m^2, the ranges of LWD and SWD were comparable for cafs_bsl and cafs_ini, and likely the radiative fluxes are dominated by the persistent low-level liquid bearing clouds common to both simulations (Figures 8g and 8h) However, when LWPs are smaller, cafs_ini shows a greater decline in the amount of incoming SWD, that is, larger cloud attenuation. The results in Figure 10 highlight that small differences in cloud-phase partitioning can yield substantial ranges in radiative energy fluxes.

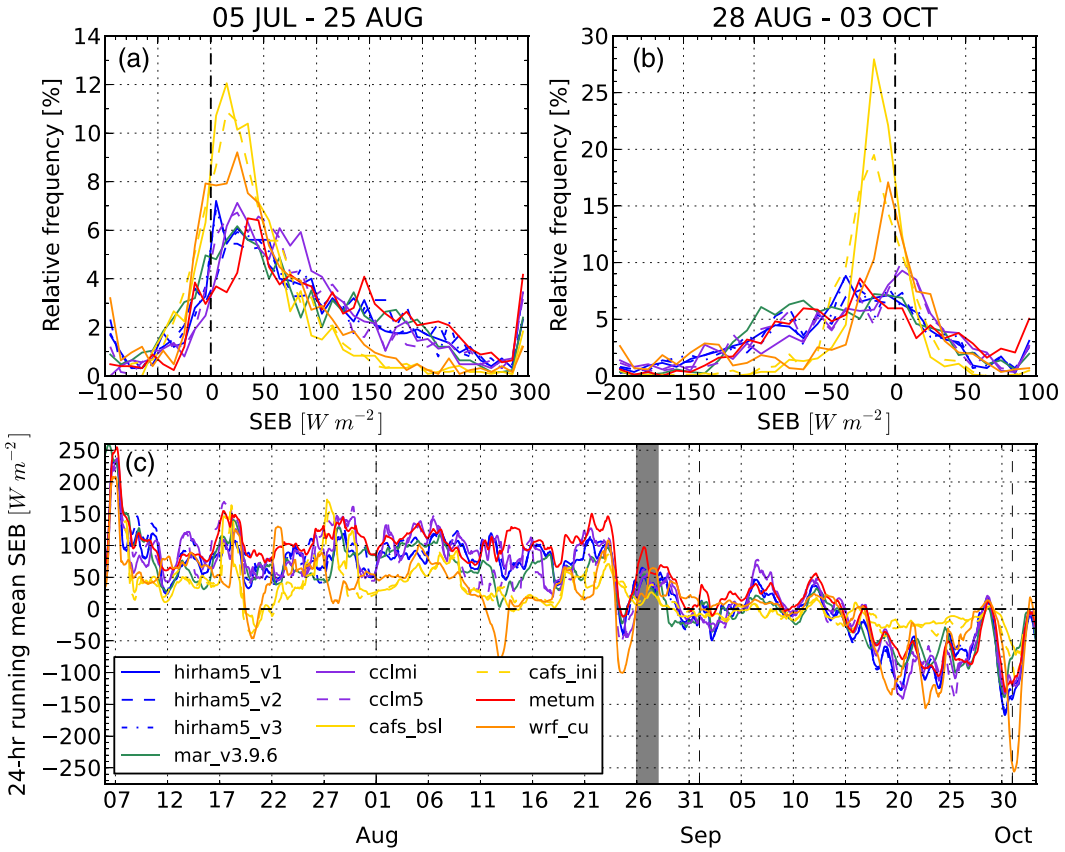

Figure 11. Relative frequency distributions of the models' full surface energy budget (SEB; W/m^2) split into two periods, (a) 5 July through 25 August and (b) 28 August through 3 October. (c) Twenty-four-hour running mean time series of models' SEB, with the 2-day separation between periods shaded dark gray. Positive SEB indicates a surplus of energy into the surface; negative indicates a deficit of energy.

3.4. Modeled Total SEB

This section focuses on how the models represented the total SEB (SWN + LWN + SHF + LHF; SWN = net shortwave radiation, LWN = net longwave radiation). We do not attempt to estimate the observed SEB for two primary reasons. First, upwelling radiation fluxes were not directly measured during ACSE. While LWN is later estimated using the skin temperature (sections 3.5 and 3.6), SWN requires accurate estimates of surface albedo; small deviations from reality, especially over 1-hr time spans when the icebreaker was moving through variable surface conditions, would have drastic effects on the estimated SWN values. Moreover, even if a local albedo could have been estimated from the web-cam monitoring, it would not be the areal average albedo over many km^2 that the models use. Second, the data recovery rate for observed turbulent fluxes (21% and 65% for the LHF and SHF, respectively) would further bias the analysis to specific times. Instead, we compare how the models' SEB evolved, as well as examining the modeled SEB statistical distributions for the two seasonal periods discussed previously.

The 24-hr running-mean time series of modeled SEB in Figure 11c confirms that the models did capture the observed seasonal transition discussed in Sotiropoulou et al. (2016). Prior to 27 August, modeled SEB distributions peaked on the positive side (energy to the surface; e.g., melt season), with long tails skewed toward large net energy surpluses and much fewer negative values (Figure 11a). After 27 August, the distribution peaks shifted toward negative SEB values with more pronounced negatively skewed tails (Figure 11b); this shift signals the onset of the surface freeze-up.

The SEB distributions simulated by cafs and wrf_cu differ from the other models. Their RFDs have more well-defined and narrower peaks for both periods. For cafs, the lower variability, especially for largely

positive SEB values in Figure 11a, is likely a combination of excessive cloud liquid shielding the surface from larger SWD (Figure 5a), and too high surface albedo from too much sea-ice cover in the fully coupled runs, causing an excessive amount SWD that did penetrate the clouds to be reflected by the surface. Furthermore, the temporal variability of cafs SEB simulations typically revealed the smallest absolute changes in time, especially during the latter period (Figure 11c). The RFDs of SWD and LWD, and the cloud emissivity (Figure 9), are representative of a too opaque atmosphere in wrf_cu. Further contributing to wrf_cu's smaller SEB is thicker sea ice specified and too large SIC relative to the most other models (Figure S8), further restricting the amount of shortwave that could be absorbed at the surface. The wrf_cu SEB also had lower residuals during the first period (occasionally <0 W/m^2) relative to the other models with prescribed sea-ice conditions (Figure 11c). The metum model consistently had the largest SEB values, consistent with its positive biases in SWD (Figure 5a), while all hirham5 and cclm simulations feature broadly similar SEBs over the entire period, consistent with their similar radiative and turbulent energy distributions (Figure 5).

The mean intermodel SEB spread was ~90 W/m^2 before 27 August and ~50 W/m^2 after (Figure 11c). However, despite the large intermodel variability, all models showed similar SEB tendencies that appear critical to the evolution of the seasonally transitioning Arctic. These include the sharp dip from a consistently positive to briefly negative residuals between 23–25 August, seemingly a regime change during the last week of August, and a return to net-zero SEB variability transition during the first 2 weeks of September. Similar circumstances involving a regime change followed by a transitional period were also observed in August–September of 2008 during the ASCOS field experiment (Sedlar et al., 2011).

3.5. Lower Tropospheric Stability and Model Processes

Clouds containing super-cooled liquid water over Arctic sea ice are a main controller of lower-troposphere stratification (Brooks et al., 2017; Sedlar et al., 2012; Sedlar & Shupe, 2014; Shupe et al., 2008; Shupe et al., 2013). Statistical differences between modeled and observed water paths and cloud-radiative interactions highlighted in Figures 7–10 indicate that important cloud-driven processes typically observed in the Arctic may be missing, or misrepresented, in the RCMs. It is therefore important to make the connections from modeled biases in clouds and radiation to potential biases in inter-connected processes.

One metric to combine analysis of atmospheric stratification with cloud cover is the LWN-LTS (lower tropospheric stability) relationship. Like in Wood and Bretherton (2006), LTS is estimated as the difference in equivalent potential temperature between two levels, here the 950 hPa pressure level and the surface. A conserved measure in a moist environment, the equivalent potential temperature difference, provides a measure of the lower-level static stability up to 950 hPa, typically 450–600 m AGL during ACSE. Hence, 950 hPa is often above the relatively shallow surface-generated turbulent boundary layer (Sotiropoulou et al., 2016; Tjernström et al., 2004; Tjernström et al., 2012) but frequently within the subcloud overturning mixed-layer (e.g., Brooks et al., 2017). This layer is frequently associated with mixed-phase stratocumulus often observed between 400 m and 2 km AGL (Sedlar et al., 2012; Shupe et al., 2008; Sotiropoulou et al., 2016; Tjernström et al., 2012); the turbulent surface-generated and cloud-driven subcloud layers can be either coupled or decoupled. Recent studies found that weak decoupling occurred most frequently (Brooks et al., 2017; Sedlar & Shupe, 2014; Shupe et al., 2013; Sotiropoulou et al., 2016). To relate to the cloud-radiative state, surface LWN is analyzed together with LTS. LWN is strongly dependent upon two parameters: the presence/absence of liquid-water clouds, generally with a large infrared emissivity, and the temperature of the surface. When combined, LWN-LTS distributions can provide useful relationships and indicate regimes combining the surface state, emissivity of the overlying atmosphere/cloud, and the degree to which these mechanisms influence stability.

Exploring the joint RFD of observed LWN and LTS gives an indication of the most common cloud-stability regimes during ACSE (Figure 12a). A well-defined distribution peak occurred at near-neutral stratification, LTS within ± 2 K and LWN between 0 and -10 W/m^2. In summer-autumn, when low-level stratocumulus contains liquid droplets over Arctic sea ice, LWN generally falls within the 0 to -10 W/m^2 range (Sedlar et al., 2011). Coupled with stratification, this peak is characteristic of intermittent coupling between surface-generated and cloud-generated turbulent mixing layers. This is an important regime in that it indicates potential communication, that is, transfer of heat and moisture, between surface and cloud.

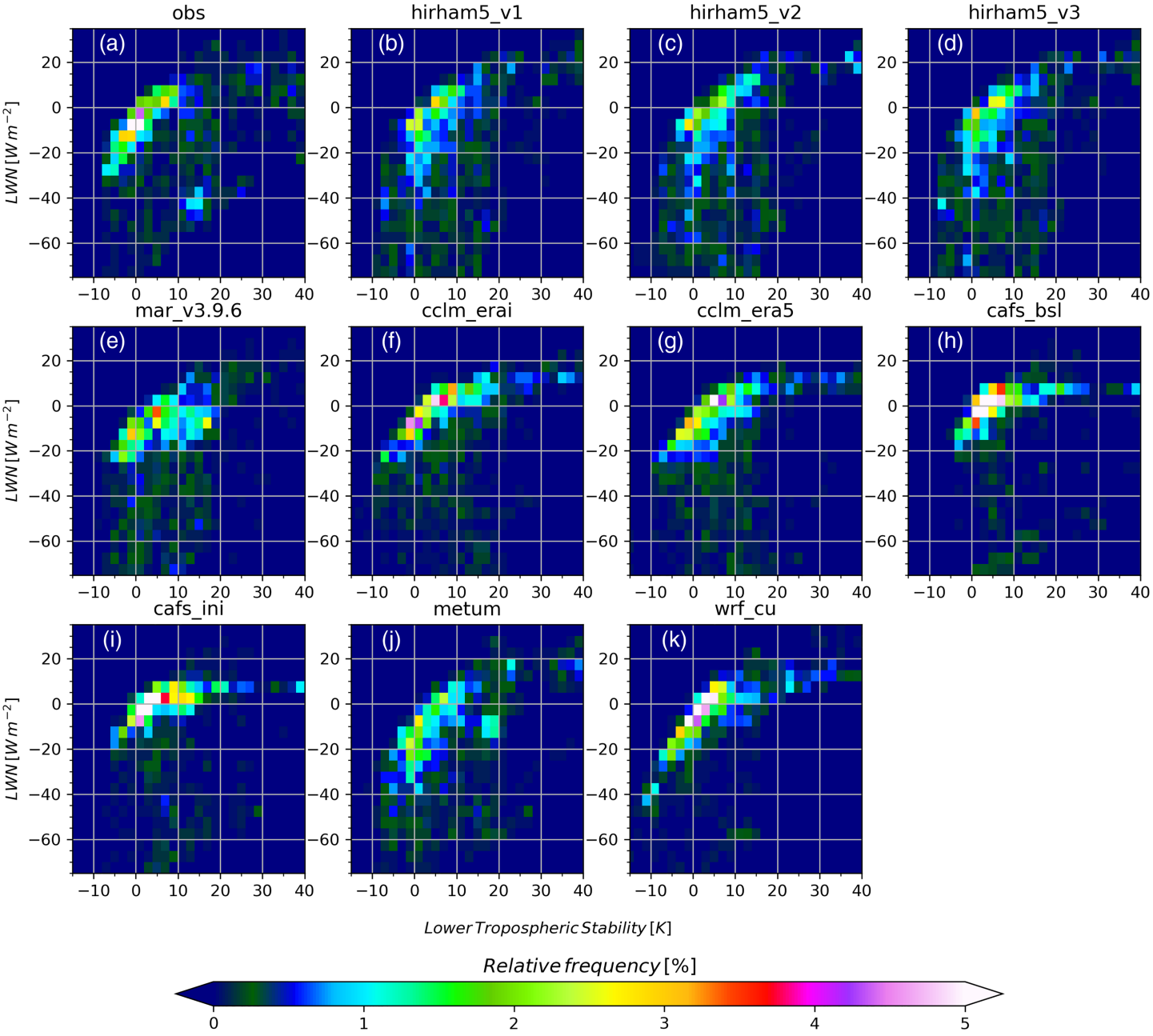

Figure 12. Relative frequency distributions showing the relationship between surface net longwave radiation (LWN; W/m^2] and lower tropospheric stability (LTS, [K]) for (a) observations, (b) hirham5_v1, (c) hirham5_v2, (d) hirham5_v3, (e) mar_v3.9.6, (f) cclmi, (g) cclm5, (h) cafs_bsl, (i) cafs_ini, (j) metum, and (k) wrf_cu. See text for details. Statistics are valid for the full Arctic Clouds in Summer Experiment (ACSE) campaign, 5 July at 01:00 UTC through 3 October at 07:00 UTC.

Around this peak, the observed distributions also indicate relatively frequent unstable stratification (LTS < 0 K) with LWN as low as −30 W/m^2. This regime is likely associated with a colder air mass over warmer, open water, supporting convection through warmer skin temperatures keeping the lower troposphere well-mixed, at the same time increasing the LWN deficit. Another observed regime was positive LTS (stable) together with positive LWN (gain for the surface). This regime tended to occur for two different synoptic situations. The first in warm-air advection connected with synoptic frontal passages, causing larger longwave radiation from the cloud than upwelling from the melting ice. This regime resulted in LTS ranging approximately 5 to 15 K. The second situation led to the extremely large, but infrequent, LTS in

connection with positive LWN radiation. This was associated with strong warm and moist air advection from the nearby continent. As this air encountered and traversed over a melting, colder sea-ice surface, air-mass modification processes developed a strong surface temperature inversion, while near-surface saturation developed fog or shallow low clouds (e.g., Tjernström et al., 2019). The combination of low clouds trapped in a stable and warm lower troposphere led to large LTS and positive LWN. In ACSE this regime was primarily associated with an 8-day event in early August (Tjernström et al., 2015), although stable, moist inversions did occur during other instances at ACSE (Tjernström et al., 2019) and climatologically occur with a modest frequency (Sedlar & Tjernström, 2017).

The final important regime is the largely negative LWN (< -40 W/m^2) associated with a radiatively clear sky in connection with intermediate LTS values, 10–20 K (Figure 12a). Lack of radiatively opaque low clouds eliminates one mechanism driving lower tropospheric mixing, that is, cloud buoyancy, and as a feedback promotes a large LWN deficit, which isolates this regime into further stable stratification.

The modeled LWN-LTS relationships, and their representation of the associated regimes in the LWN-LTS phase space, are shown in Figures 12b–12k (see Figure S14 for model-observation differences). Broadly, the modeled relationships were similar to the observations for the most common regime: the cloud-driven near-adiabatic mixing of the lower troposphere. Hence, models captured the frequent weak or intermittent coupling between the surface and 950-hPa level, likely crucial in maintaining the heat and moisture exchange between the surface and lower clouds. However, most models underestimated the occurrence of the well-mixed (LTS < 0 K) regime (Figure S14). Instead, they favored a slightly more stable regime with moderate ranges of LWN deficit (-10 to -30 W/m^2). The fully coupled surface-atmosphere cafs runs corroborate the close connection between surface conditions and LWN-LTS relationships; cafs distributions were the least variable and had the least occurrence of unstable stratification (Figures 12h and 12i and S14), a consequence of a surface predominantly modeled as ice-covered.

Another common discrepancy among all models was a lack of the observed clear-sky, stable regime. The hirham5 (b–d), mar_v3.9.6 (e), cclm (f and g), and metum (j) simulations all showed the presence of clear-sky conditions (LWN < -40 W/m^2), but the observed moderate stability associated with this regime (10–20 K) was often replaced with modest stability (0–10 K), indicating an apparent disconnect between decoupling and surface-based stability. Both cafs (h and i) and wrf_cu (k) generally failed to produce any clear-sky stable scenes, consistent with overestimated LWPs, frequent blackbody liquid clouds, and the associated radiative-flux biases shown in Figures 7–10. Instances of stably stratified clear-sky events have been characterized as crucially important in dictating the transition between seasons over Arctic sea ice, from melt to onset of freeze (Persson, 2012; Sedlar et al., 2011; Sotiropoulou et al., 2016). The differences between observations and models in the occurrence and radiative impact of this cloud-stability regime signify a critical shortcoming in modeling a fundamental Arctic atmospheric process.

Additionally concerning is the models' representation of moderate to large positive LTS associated with a surplus in LWN. The models overrepresented this regime, especially for the more moderate stability range, LTS ~10–25 K (Figure S14), compared to the observations. This regime was primarily associated with significant horizontal advection and air mass transformation, resulting in anomalously large and positive LWN (Tjernström et al., 2015). The differences in LWN-LTS distributions between the models and observations, and among the models themselves, indicate that these types of large-scale advective events pose complications for the models; an example of this is explored in the next section.

3.6. Warm, Moist Advective Event

The ACSE campaign directly observed an intense warm and moist advection event across the lower troposphere to midtroposphere, between 30 July and 8 August 2014 (see Figures S4 and S5). This event led to anomalous sea-ice melt (Tjernström et al., 2015) primarily through anomalies of large, positive LWN and SHF into the ice surface, both of which were driven by low clouds trapped within a very warm inversion associated with advection (Tjernström et al., 2015). To examine this important event, statistical ranges of LWN were calculated for three distinct 7-day periods characterizing its evolution: (1) preevent [23–30 July], (2) core event [31 July to 6 August], and (3) postevent [8–14 August].

The 10th–90th and interquartile ranges of LWN revealed substantial transitions between the three periods. Observed median LWN jumped from -20 to 5 W/m^2 from preevent to core event (Figure 13, black); the

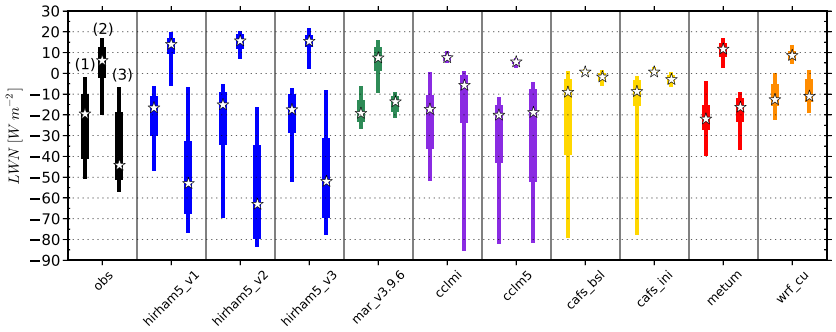

Figure 13. Observed (black) and modeled (colors) percentile ranges of net longwave radiation (LWN; W/m^2] for three 1-week periods around the warm, moist advection event in early August; period (1): 23–30 July; period (2): 31 July to 6 August; and period (3): 8–14 August. The thin bars are the 10th–90th percentile bounds, the thick bars are the 25th–75th percentile bounds, and the white stars are the medians, for each week-long period.

median transition was even larger into the postevent week, from 5 to −45 W/m^2. These swings in LWN are consistent with the advection of warm, moist air trapping a low, liquid-bearing cloud layer, followed by a drying and subsequent low-cloud dissipation.

The simulations also revealed distinct transitions between the three periods, although to various degrees. The transitions most similar to the observations were found in hirham5 simulations. However, the magnitudes of LWN, both during the core event and postevent, were considerably larger than observed; the larger positive LWN values during period (2) are consistent with the overestimated occurrence frequency of positive LWN in connection with the LTS metric (Figures 12b–12d). While all the remaining models similarly resolved a LWN surplus during the core event, the variability of LWN during this week was considerably smaller than observed. The 10th–90th percentile ranges of LWN for cclm, cafs, and wrf_cu simulations were nearly indistinguishable from the median values, revealing a persistent, positive energy input on the surface. But potentially more critical to the processes driving the surface energy balance is that all models besides hirham5 did not transition to large LWN deficits, reflective of clear or intermittently cloudy skies, during the postevent period (3) as indicated in the observations. Instead, mar_v3.9.6, cclmi, cafs, metum, and wrf_cu had LWN medians and interquartile ranges resulting in smaller LWN deficits during period (3) than during period (1). This suggests that the modeled clouds in response to the advective event, and after its passing, failed to represent the appropriate mechanisms contributing to cloud formation and dissipation.

4. Discussion

This suite of six state-of-the-art RCMs, providing 10 total simulations of the ACSE field campaign, permitted a detailed evaluation regarding the models' capabilities in simulating observed atmospheric and SEB properties and processes over the Arctic. Here, we discuss some of the major findings in the context of how model-specific processes and energy inputs may be governing the sensitive evolution of the SEB.

The most significant biases between models and observations emerge from the model physics, as all the models use the same forcing (ERA-Interim). The radiative fluxes represent the largest energy input, and consequently, their absolute biases are also the largest. While the hirham5 and cclm agreed reasonably with the observed radiation and LWP distributions, the other models could be grouped by their biases in downwelling radiation. Both cafs and wrf_cu underestimated SWD while exhibiting too little variability in LWD; unsurprisingly, these models also overestimated the amount of cloud liquid, especially low liquid-bearing clouds. Increasing the IN concentration in cafs_ini did little to reduce cloud LWPs, except for a reduction in cloud liquid at pressure levels lower than 800 hPa, because the lower atmosphere was above freezing for the majority of the first 2 months; there were also marginal increases in IWP between the two cafs simulations even during the second colder period, suggesting that the increased IN concentration resulted in ice crystal precipitation and deposition. The mar_v3.9.6 and metum overestimated SWD and underestimated LWD; these models underrepresented the amount of cloud liquid. Interestingly, metum's colder skin temperature bias (Figure 4b) is consistent with a lower frequency of liquid-bearing clouds (Sedlar et al., 2011). However,

skin temperature bias for the mar_v3.9.6 was very similar to the other simulations with larger LWP. Therefore, consistent underestimation of LWD in this model does not appear to be driving the cold bias at the surface, suggesting some error compensations in mar_v3.9.6, as already highlighted over Greenland by Fettweis et al. (2017).

With turbulent fluxes, the models captured the qualitative distribution shape of the observations, but modest distribution differences are intriguing. The cafs simulations displayed a dominant presence of negative SHFs (cooling the surface), while the remaining models and observations peaked at weakly positive fluxes. The fully coupled cafs simulations also considerably overestimated the sea-ice cover, further causing the surface temperatures to be too cold. For negative SHFs to dominate, air temperatures above the surface must also be too cold, a result that was confirmed by the analysis of temperature profiles. Therefore, for cafs, it is the atmospheric thermodynamic structure, and not an overly warm surface, that was driving the turbulent heat-flux distributions. At the same time, we found negative SHFs were common for cafs even when temperature gradients suggest a stable stratification in the surface layer (Figures 6g and 6h); we attribute these counter-gradient fluxes to the influence of open water (leads, large melt ponds) in the otherwise large sea-ice concentrations. An even larger cold bias was present in wrf_cu's lower troposphere; this model also had large negative SHF distribution tails, consistent with atmospheric biases guiding its near-surface turbulent heat biases. The covariability of LWN and LTS indicated that both the cafs and wrf_cu simulations contained the most excessive static mixing regime associated with small deficits in LWN, without the occurrence of the stably stratified and radiatively clear regime that was present in the observations and other models. These distributions were indicative of an over-abundance of cloud driven mixing, supporting turbulent heat transfer away from the surface to the atmosphere. These findings support a coupled surface-lower atmosphere structure that can actively exchange momentum, heat, and moisture and potentially sustain the abundance of low liquid clouds so prevalent in these models. Relatively large positive biases in SHF were observed in the cclm simulations. When scaling the SHF by 10-m wind speed, it was found that the overly excessive SHFs during stable stratification are likely connected to errors in the exchange coefficient of sensible heat within the model's surface flux parameterization scheme (Figures 6e and 6f).

Differences in radiation and turbulent heat flux distributions are directly responsible for the differences in total SEB distributions among the models. Despite critical energy input biases, the models agreed well on the evolution of the seasonal transition of the SEB. This must be intimately connected to daily (or subdaily) specification of SIC and sea surface temperature, and the application of large-scale atmospheric nudging or daily/subdaily model forecast initialization. Even while sufficiently simulating the synoptic environment, the LWN transitions surrounding the extreme warm, moist advective event clearly were not in agreement among the models, or between the models and observations. The critical processes in connection with this event, the strong surface-based inversion and trapping of low clouds, were not universal among the models. Furthermore, at any given time, the spread in SEB among models during the first 2 months of ACSE was nearly 100 W/m^2; this spread decreased to around 50 W/m^2 through September and early October. Such a spread in SEB residuals is large, and it is obvious that each model's SEB solution cannot represent what occurred in reality and would also give vastly different results if the models would have been coupled to a sea-ice and ocean model.

Having at least one fully coupled model, cafs, provided a unique opportunity to assess how its SEB evolved when the open water, sea ice, snow, and melt pond surfaces responded to biases in energy inputs. Too large liquid-cloud presence shielded the surface too much from shortwave radiation. Subsequently, the largest SEB residuals (those greater than 125 W/m^2) were severely underestimated. Without this heat being absorbed at the surface, especially during July with the sun at a higher elevation, sea-ice coverage is sustained, during times and locations where the other models, forced by observed SIC, suggest the open water fraction (leads, melt ponds) should have been larger; overactive turbulent mixing near the surface further adds to the cooling at the surface. As a result, the SEB time series for both cafs simulations were among the lowest in absolute magnitude. However, these SEB evolutions were also the most physically consistent with the excessive liquid clouds and their impact on the radiative and turbulent fluxes.

Oppositely, the metum and cclm simulations generally showed the largest absolute SEB magnitudes but were connected to different processes. For metum, the lack of relatively large LWP clouds permitted too much shortwave radiation to be absorbed by surfaces that generally had smaller SICs. Had the metum (or

any other atmosphere-only model) been fully coupled, it is unclear how the general lack of liquid clouds, and subsequent overestimation of SWD, would have impacted its energy budget evolution. SICs for cclm typically matched well with the satellite observations (Figures S8e and S8f), but the LWD, in conjunction with more liquid-bearing clouds, led to a narrower distribution relative to the observations and other RCMs. There was also a shift to more positive SHFs for cclm, further transferring energy toward the surface, a bias that has been attributed to the behavior of the surface flux parameterization scheme under statically stable conditions. Taken together, these biases caused the cclmi and cclm5 SEBs to be consistently large, especially prior to 1 September. The statistical comparisons between cclmi and cclm5 showed very little differences, suggesting only modest differences in conjunction with the lateral boundary forcing between the ERA-5 and ERA-Interim reanalyses.

Improved understanding of critical, yet infrequent, events is a necessary future step in the analysis of RCM simulations. However, a more fundamental step is needed to address the variability related to cloud-radiative interactions. By only increasing the IN concentrations, the cafs_ini simulation revealed a better agreement with observed cloud LWP compared to the baseline simulation. Although we found that while the lower atmosphere IWC increased marginally, the LWCs were statistically as large as for the baseline simulation; instead the increased IN lead to a reduction in liquid-bearing clouds higher up in the free troposphere, resulting in a smaller, total column-integrated LWP. However, low-level liquid clouds exert the largest impact on surface radiation, and therefore, we found very little difference in SWD and LWD between cafs simulations. The hirham5 simulations employing two different cloud schemes led to only modest differences in the LWP and IWP distributions; however, the cloud-radiative interactions associated with these LWP differences did impact the downwelling radiative fluxes. In particular, we found the simulations applying the Sundqvist et al. (1989) microphysical parameterization failed to retain supercooled liquid and therefore were unable to sustain low-level clouds emitting as blackbodies, while the Tompkins (2002) parameterization did (Figures 9a–9c). At the same time, clouds with emissivity below 1 (graybody clouds) were more often associated with a somewhat colder emission temperature with the Tompkins versus the Sundqvist et al. scheme, and therefore, differences in LWD emerged among the hirham5 simulations, especially for smaller LWPs that are responsible for the graybody clouds (Figure 10). This seesaw-type behavior highlights the struggle present when evaluating these models.

5. Conclusions

To conclude, the following points represent the main conclusions from this study:

- Different reanalysis forcing (ERA-Interim and ERA5) the lateral boundaries and nudging for cclm simulations had minimal impact on the physical processes determining the surface energy fluxes and the cloud properties. Whether or not this is a universal conclusion applicable to all RCMs employing different parameterization schemes is unknown and deserves attention.
- Variation among the simulated cloud properties is the most important factor in correctly representing the surface energy fluxes. Errors in the amount of cloud condensate, especially the cloud liquid, directly influenced the radiative fluxes, and indirectly modified the turbulent heat fluxes through modification of the near-surface mixing state. RCMs continue to struggle with these important cloud forcing processes over the Arctic.
- In this respect, we conclude that the situation has not improved vastly since the ARCMIP simulations over a decade ago.
- The importance of longwave cloud radiative forcing in the Arctic is well established, and we found that it is critical in determining the static stability of the lower troposphere. However, during the middle to late summer months, accurate representation of the incoming shortwave radiation is also found to be crucial. Shortwave radiation typically is the largest energy source to the surface, and we found that models that overestimate the SWD had the largest SEB residuals, while those with too little shortwave had the lowest residuals. Again, properly describing the cloud properties is of utmost importance, and unfortunately, here is where the state-of-the-art RCMs continue to underperform.

Regardless of the cause, these distributions and errors in clouds and radiation are still similar to those reported by studies evaluating RCMs that simulated the SHEBA campaign from years past (Inoue et al., 2006; Rinke et al., 2006; Tjernström et al., 2005, 2008; Wyser et al., 2008). Undoubtedly, the

scientific understanding of the processes underlying the modeling problems reported on a decade and a half ago is much better today, but still it seems only that marginal improvements have been realized in the models. We will offer only one recommendation: that model development is performed both in atmosphere-alone (or ice-ocean-alone) framework and in a fully coupled framework. This puts additional constraints on both the observational and modeling science communities, since both the modeling and the observations need to consider the coupling between different systems. This is timelier now than ever before, with the MOSAiC field campaign set to provide a full year of process observations adapted to a truly coupled framework for the observations. This will provide a unique dataset optimally suitable for a new set of RCM intercomparisons, but this time with several fully coupled modeling systems; this is highly recommended.

Acknowledgments
We thank the ECMWF for making reanalysis products freely available. The CCLM contribution (G. H.) was funded by the Federal Ministry of Education and Research (BMBF) under grant 03F0776D (CATS). A. R. acknowledges the funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project 268020496—TRR 172, within the Transregional Collaborative Research Center "ArctiC Amplification: Climate Relevant Atmospheric and SurfaCe Processes, and Feedback Mechanisms (AC)^3. A. S. acknowledges support from NOAA/ESRL Physical Science Division in supporting CAFS. M. S. acknowledges that the wrf_cu effort was supported by the Regional and Global Model Analysis (RGMA) component of the U.S. Department of Energy's Office of Science, as contribution to the HiLAT-RASM project. The observational and regional climate model data sets used in this study have been merged into two NetCDF files and are published and freely available from the following web address: https://bolin.su.se/data/swerus-2014-rcm-metobs (Sedlar, 2020).

References

Akperov, M., Rinke, A., Mokhov, I. I., Matthes, H., Semenov, V. A., Adakudlu, M., et al. (2018). Cyclone activity in the Arctic from an ensemble of regional climate models (Arctic CORDEX). *Journal of Geophysical Research: Atmospheres*, *123*, 2537–2554. https://doi.org/10.1002/2017JD027703

AMAP: Snow, Water, Ice and Permafrost in the Arctic (SWIPA) 2017. (2017), Arctic Monitoring and Assessment Programme (AMAP), Oslo, Norway, xiv + 269 pp.

Andreas, E. L., Persson, P. O. G., Jordan, R. E., Horst, T. W., Guest, P. S., Grachev, A. A., & Fairall, C. W. (2010). Parameterizing turbulent exchange over sea ice in winter. *Journal of Hydrometeorology*, *11*, 87–104. https://doi.org/10.1175/2009JHM1102.1

Berg, P., Döscher, R., & Koenigk, T. (2013). Impacts of using spectral nudging on regional climate model RCA4 simulations of the Arctic. *Geoscientific Model Development*, *6*, 849–859. https://doi.org/10.5194/gmd-6-849-2013

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. . L. H., Ménard, C. B., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description—Part 1: Energy and water fluxes. *Geoscientific Model Development*, *4*(3), 677–699. https://doi.org/10.5194/gmd-4-677-2011

Brinkop, S., & Roeckner, E. (1995). Sensitivity of a general circulation model to parameterizations of cloud-turbulence interactions in the atmospheric boundary layer. *Tellus A*, *47*, 197–220.

Brooks, I. M., Tjernström, M., Persson, P. O. G., Shupe, M. D., Atkinson, R. A., Canut, G., et al. (2017). The turbulent structure of the Arctic summer boundary layer during the Arctic summer cloud-ocean study. *Journal of Geophysical Research: Atmospheres*, *122*, 9685–9704. https://doi.org/10.1002/2017JD027234

Cassano, J. J., DuVivier, A., Roberts, A., Hughes, M., Seefeldt, M., Brunke, M., et al. (2017). Development of the Regional Arctic System Model (RASM): near-surface atmospheric climate sensitivity. *Journal of Climate*, *30*(15), 5729–5753. https://doi.org/10.1175/JCLI-D-15-0775.1

Christensen, O. B., Drews, M., Christensen, J. H., Dethloff, K., Ketelsen, K., Hebestadt, I., & Rinke, A. (2007). *The HIRHAM Regional Climate Model Version 5 (b); DMI technical report 06–17* (p. 22). Copenhagen, Denmark: Danish Meteorological Institute.

Curry, J. A., & Lynch, A. H. (2002). Comparing Arctic regional climate model. *Eos*, *83*, 87.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Doms, G., Förstner, J., Heise, E., Herzog, H. J., Rahsendorfer, M., Reinhardt, T., et al. (2011). A description of the non-hydrostatic regional model LM. Part II: Physical parameterization, core documentation. Consortium for small-scale modelling. http://www.cosmo-model.org.

Duynkerke, P. G. (1988). Application of the E-turbulence closure model to the neutral and stable atmospheric boundary layer. *Journal of the Atmospheric Sciences*, *45*(5), 865–880.

Fettweis, X., Box, J. E., Agosta, C., Amory, C., Kittel, C., Lang, C., et al. (2017). Reconstructions of the 1900-2015 Greenland ice sheet surface mass balance using the regional climate MAR model. *The Cryosphere*, *11*, 1015–1033. https://doi.org/10.5194/tc-11-1015-2017

Furtado, K., & Field, P. (2017). The role of ice microphysics parameterizations in determining the prevalence of supercooled liquid water in high-resolution simulations of a southern ocean midlatitude cyclone. *Journal of the Atmospheric Sciences*, *75*, 2001–2021. https://doi.org/10.1175/JAS-D-16-0165.1

Gallèe, H., Guyomarch, G., & Brun, E. (2001). Impact of snow drift on the Antarctic ice sheet surface mass balance: Possible sensitivity to snow-surface properties. *Boundary-Layer Meteorology*, *99*, 1. https://doi.org/10.1023/A:1018776422809

Garrett, T. J., Radke, L. F., & Hobbs, P. V. (2002). Aerosol effects on cloud emissivity and surface longwave heating in the Arctic. *Journal of the Atmospheric Sciences*, *59*, 769–778.

Garrett, T. J., & Zhao, C. (2006). Increased Arctic cloud longwave emissivity associated with pollution from mid-latitudes. *Nature*, *440*(7085), 787–789. https://doi.org/10.1038/nature04636

Giorgi, F., Jones, C., & Asrar, G. R. (2009). Addressing climate information needs at the regional level: The CORDEX framework. *WMO Bulletin*, *58*(3), 175–183.

Gutjahr, O., & Heinemann, G. (2018). A model-based comparison of extreme winds in the Arctic and around Greenland. *International Journal of Climatology*, *38*, 5272–5292. https://doi.org/10.1002/joc.5729

Hersbach, H. et al. (2018), Operational global reanalysis: progress, future directions and synergies with NWP, *ERA Report Series*, https://doi.org/10.21957/tkic6g3wm.

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*, 2318–2341.

Illingworth, A. J., Hogan, R. J., O'Connor, E. J., Bouniol, D., Brooks, M. E., Delanoé, J., et al. (2007). Cloudnet: Continuous evaluation of cloud profiles din seven operational models using ground-based observations. *Bulletin of the American Meteorological Society*, *88*(6), 883–898. https://doi.org/10.1175/BAMS-88-6-883

Inoue, J., Liu, J., Pinto, J. O., & Curry, J. A. (2006). Intercomparison of Arctic regional climate models: Modeling clouds and radiation for SHEBA in may 1998. *Journal of Climate*, *19*, 4167–4178.

Intrieri, J. M., Fairall, C. W., Shupe, M. D., Persson, P. O. G., Andreas, E. L., Guest, P. S., & Moritz, R. E. (2002). An annual cycle of Arctic surface cloud forcing at SHEBA. *Journal of Geophysical Research*, *107*(C10), 8039. https://doi.org/10.1029/2000JC000439

Jiménez, P. A., Dudhia, J., Fidel Gonzàlez-Rouco, J., Navarro, J., Montàvez, J. P., & Garcìa-Bustamante, E. (2012). A revised scheme for the WRF surface layer formulation. *Monthly Weather Review*, *140*, 898–918. https://doi.org/10.1175/MWR-D-11-00056.1

Klaus, D., Dethloff, K., Dorn, W., Rinke, A., & Wu, D. L. (2016). New insight of Arctic cloud parameterization from regional climate model simulations, satellite-based, and drifting station data. *Geophysical Research Letters*, *43*, 5450–5459. https://doi.org/10.1002/2015GL067530

Klein, S. A., McCoy, R. B., Morrison, H., Ackerman, A. S., Avramov, A., Boer, G. ., et al. (2009). Intercomparison of model simulations of mixed-phase clouds observed during the ARM Mixed-Phase Arctic Cloud Experiment. I: Single-layer cloud. *Quarterly Journal of the Royal Meteorological Society*, *135*(641), 979–1002. https://doi.org/10.1002/qj.416

Kohnemann, S., Heinemann, G., Bromwich, D., & Gutjahr, O. (2017). Extreme warming in the Kara Sea and Barents Sea during the winter period 2000 to 2016. *Journal of Climate*, *30*, 8913–8927. https://doi.org/10.1175/JCLI-D-16-0693.1

Lock, A. P., Brown, A. R., Bush, M. R., Martin, G. M., & Smith, R. N. B. (2000). A new boundary-layer mixing scheme. 1. Scheme description and single-column model tests. *Monthly Weather Review*, *128*, 3187–3199.

Louis, J.-F. (1979). A parametric model of vertical eddy fluxes in the atmosphere. *Boundary-Layer Meteorology*, *17*, 187–202.

Lubin, D., & Vogelmann, A. M. (2006). A climatologically significant aerosol longwave indirect effect in the Arctic. *Nature*, *439*(7075), 453–456. https://doi.org/10.1038/nature04449

Mellor, G. L., & Yamada, T. (1974). A hierarchy of turbulence closure models for planetary boundary layers. *Journal of the Atmospheric Sciences*, *31*, 1791–1806.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*(D14), 16,663–16,682.

Moat, B. I., Yelland, M. J., & Brooks, I. M. (2015). Airflow distortion at instrument sites on the ODEN during the ACSE project, National Oceanography Centre, 114 pp., (National Oceanography Centre Internal Document, 17), Southampton, GB.

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Monthly Weather Review*, *137*, 991–1007. https://doi.org/10.1175/2008MWR2556.1

Nakanishi, M., & Niino, H. (2006). An improved Mellor-Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Boundary-Layer Meteorology*, *119*, 397–407. https://doi.org/10.1007/s10546-005-9030-8

Orr, A., Phillips, T., Webster, S., Elvidge, A., Weeks, M., Hosking, S., & Turner, J. (2014). Met Office Unified Model high-resolution simulations of a strong wind event in Antarctica. *Quarterly Journal of the Royal Meteorological Society*, *140*, 2287–2297. https://doi.org/10.1002/qj.2296

Persson, P. O. G. (2012). Onset and end of the summer melt season over sea ice: Thermal structure and surface energy perspective from SHEBA. *Climate Dynamics*, *39*(6), 1349–1371. https://doi.org/10.1007/s00382-011-1196-9

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), 8045. https://doi.org/10.1029/2000JC000705

Persson, P. O. G., Shupe, M. D., Perovich, D., & Solomon, A. (2017). Linking atmospheric synoptic transport, cloud phase, surface energy fluxes, and sea-ice growth: Observations of midwinter SHEBA conditions. *Climate Dynamics*, *49*(4), 1341–1364. https://doi.org/10.1007/s00382-016-3383-1

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., et al. (2017). The Weather Research and Forecasting model. Overview, system efforts and future directions. *Bulletin of the American Meteorological Society*, *98*(8), 1717–1737. https://doi.org/10.1175/BAMS-D-15-00308.1

Prytherch, J., Yelland, M. J., Brooks, I. M., Tupman, D. J., Pascal, R. W., Moat, B. I., & Norris, S. J. (2015). Motion-correlated flow distortion and wave-induced biases in air-sea flux measurements from ships. *Atmospheric Chemistry and Physics*, *15*, 10,619–10,629. https://doi.org/10.5194/acp-15-10619-2015

Rinke, A., Dethloff, K., Cassano, J. J., Christensen, J. H., Curry, J. A., Du, P., et al. (2006). Evaluation of an ensemble of Arctic regional climate models: Spatiotemporal fields during the SHEBA year. *Climate Dynamics*, *26*(5), 459–472. https://doi.org/10.1007/s00382-005-0095-3

Rinke, A., Lynch, A. H., & Dethloff, K. (2000). Intercomparison of Arctic regional climate simulations: Case studies of January and June 1990. *Journal of Geophysical Research*, *105*(D24), 29,669–29,683.

Ritter, B., & Geleyn, J.-F. (1992). A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Monthly Weather Review*, *120*, 303–325.

Schröder, D., Heinemann, G., & Willmes, S. (2011). The impact of a thermodynamic sea-cie module in the COSMO numerical weather prediction module on simulations for the Laptev Sea, Siberian Arctic. *Polar Research*, *30*(6334).

Sedlar, J. (2020). Meteorological observations during the SWERUS-C3 Arctic Ocean expedition 2014 and colocated regional climate model output. Dataset version 1.0. Bolin Centre database. https://doi.org/10.17043/swerus-2014-rcm-metobs

Sedlar, J., & Shupe, M. D. (2014). Characteristic nature of vertical motions observed in Arctic mixed-phase stratocumulus. *Atmospheric Chemistry and Physics*, *14*, 3461–3478. https://doi.org/10.5194/acp-14-3461-2014

Sedlar, J., Shupe, M. D., & Tjernström, M. (2012). On the relationship between thermodynamic structure and cloud top, and its climate significance in the Arctic. *Journal of Climate*, *25*, 2374–2393. https://doi.org/10.1175/JCLI-D-11-00186.1

Sedlar, J., & Tjernström, M. (2017). Clouds, warm air, and a climate cooling signal over the summer Arctic. *Geophysical Research Letters*, *44*, 1095–1103. https://doi.org/10.1002/2016GL071959

Sedlar, J., Tjernström, M., Mauritsen, T., Shupe, M. D., Brooks, I. M., Persson, P. O. G., et al. (2011). A transitioning Arctic surface energy budget: The impacts of solar zenith angle, surface albedo and cloud radiative forcing. *Climate Dynamics*, *37*(7-8), 1643–1660. https://doi.org/10.1007/s00382-010-0937-5

Shupe, M. D., & Intrieri, J. M. (2004). Cloud radiative forcing of the Arctic surface: The influence of cloud properties, surface albedo, and solar zenith angle. *Journal of Climate*, *17*, 616–628.

Shupe, M. D., Kollias, P., Matrosov, S. Y., & Schneider, T. L. (2004). Deriving mixed-phase cloud properties from Doppler radar spectra. *Journal of Atmospheric and Oceanic Technology*, *21*, 660–670.

Shupe, M. D., Kollias, P., Persson, P. O. G., & McFarquhar, G. M. (2008). Vertical motions in Arctic mixed-phase stratiform clouds. *Journal of the Atmospheric Sciences*, *65*, 1304–1322. https://doi.org/10.1175/2007JAS2479.1

Shupe, M. D., Persson, P. O. G., Brooks, I. M., Tjernström, M., Sedlar, J., Mauritsen, T., et al. (2013). Cloud and boundary layer interactions over the Arctic sea ice in late summer. *Atmospheric Chemistry and Physics*, *13*, 9379–9400. https://doi.org/10.5194/acp-13-9379-2013

Skamarock, W. C.,, Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X-Y., Wang, W., & Powers, J. G. (2008). A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-4751STR (113 pp.). https://doi.org/10.5065/D68S4MVH

Sommerfeld, A., Nikiema, O., Rinke, A., Dethloff, K., & Laprise, R. (2015). Arctic budget study of inter-member variability using HIRHAM5 ensemble simulation. *Journal of Geophysical Research: Atmospheres*, *120*, 9390–9407. https://doi.org/10.1002/2015JD023153

Sotiropoulou, G., Sedlar, J., Tjernström, M., Shupe, M. D., Brooks, I. M., & Persson, P. O. G. (2014). The thermodynamic structure of summer Arctic stratocumulus and the dynamic coupling to the surface. *Atmospheric Chemistry and Physics*, *14*, 12,573–12,592. https://doi.org/10.5194/acp-14-12573-2014

Sotiropoulou, G., Tjernström, M., Sedlar, J., Achtert, P., Brooks, B. J., Brooks, I. M., et al. (2016). Atmospheric Conditions during the Arctic Clouds in Summer Experiment (ACSE): Contrasting open water and sea ice surfaces during melt and freeze-up seasons. *Journal of Climate*, *29*(24), 8721–8744. https://doi.org/10.1175/JCLI-D-16-0211.1

Stephens, G. L. (1978). Radiation profiles in extended water clouds. II: Parameterization schemes. *Journal of the Atmospheric Sciences*, *35*, 2123–2132.

Sundqvist, H., Berge, E., & Kristjánsson, J. E. (1989). Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Monthly Weather Review*, *117*, 1641–1657.

Tjernström, M., Birch, C. E., Brooks, I. M., Shupe, M. D., Persson, P. O. G., Sedlar, J., et al. (2012). Meteorological conditions in the central Arctic summer during the Arctic Summer Cloud Ocean Study (ASCOS). *Atmospheric Chemistry and Physics*, *12*, 6863–6889. https://doi.org/10.5194/acp-12-6863-2012

Tjernström, M., Leck, C., Persson, P. O. G., Jensen, M. L., Oncley, S. P., & Targino, A. (2004). The summertime Arctic atmosphere. Meteorological measurements during the Arctic Ocean Experiment 2001. *Bulletin of the American Meteorological Society*, *85*, 1305–1322. https://doi.org/10.1175/BAMS-85-9-1305

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*, *47*, 2405–2422. https://doi.org/10.1175/2008JAMC1845.1

Tjernström, M., Shupe, M. D., Brooks, I. M., Achtert, P., Prytherch, J., & Sedlar, J. (2019). Arctic summer airmass transformation, surface inversions, and the surface energy budget. *Journal of Climate*, *32*, 769–789. https://doi.org/10.1175/JCLI-D-0216.1

Tjernström, M., Shupe, M. D., Brooks, I. M., Persson, P. O. G., Prytherch, J., Salisbury, D. J., et al. (2015). Warm-air advection, air mass transformation and fog causes rapid ice melt. *Geophysical Research Letters*, *42*(13), 5594–5602. https://doi.org/10.1002/2015GL064373

Tjernström, M., Žagar, M., Svensson, G., Cassano, J. J., Pfeifer, S., Rinke, A., et al. (2005). Modelling the Arctic boundary layer: An evaluation of six ARCMIP regional-scale models using data from the SHEBA project. *Boundary-Layer Meteorology*, *117*(2), 337–381. https://doi.org/10.1007/s10546-004-7954-z

Tompkins, A. M. (2002). A prognostic parameterization for the subgrid-scale variability of water vapor and clouds in large-scale models and its use to diagnose cloud cover. *Journal of the Atmospheric Sciences*, *59*, 1917–1942.

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), 255–275. https://doi.org/10.1175/1520-0477(2002)083<0255:SHBOTA>2.3.CO;2

Vaughan, D. G., et al. (2013). Observations: Cryosphere. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Chap. 4, pp. 317–382). Cambridge, UK and New York, NY: Cambridge University Press.

Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S., et al. (2017). The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geoscientific Model Development*, *10*(4), 1487–1520. https://doi.org/10.5194/gmd-10-1487-2017

Westwater, E. R., Han, Y., Shupe, M. D., & Matrosov, S. Y. (2001). Analysis of integrated cloud liquid and precipitable water vapor retrievals from microwave radiometers during the surface heat budget of the Arctic Ocean Project. *Journal of Geophysical Research*, *106*(D23), 32,019–32,030.

Wilson, D. R., & Ballard, S. P. (1999). A microphyiscally based precipitation scheme for the UK Meteorological Office Unified Model. *Quarterly Journal of the Royal Meteorological Society*, *125*, 1607–1639.

Wilson, D. R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D., & Morcrette, C. J. (2008). PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. *Quarterly Journal of the Royal Meteorological Society*, *134*, 2093–2107. https://doi.org/10.1002/qj.333

Wood, R., & Bretherton, C. S. (2006). On the relationship between stratiform low cloud cover and lower-tropospheric stability. *Journal of Climate*, *19*, 6425–6432.

Wyser, K., Jones, C. G., du, P., Girard, E., Willén, U., Cassano, J., et al. (2008). An evaluation of Arctic cloud and radiation processes during the SHEBA year: simulation results from eight Arctic regional climate models. *Climate Dynamics*, *30*(2-3), 203–223. https://doi.org/10.1007/s00382-007-0286-1

Zhang, J., & Rothrock, D. A. (2003). Modeling Global Sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Monthly Weather Review*, *131*, 845–861.