Evaluating Arctic clouds modelled with the Unified Model and Integrated Forecasting System

Gillian Young[1], Jutta Vüllers[1], Peggy Achtert[2], Paul Field[1,3], Jonathan J. Day[4], Richard Forbes[4], Ruth Price[1], Ewan O'Connor[5], Michael Tjernström[6], John Prytherch[6], Ryan Neely III[1,7], and Ian M. Brooks[1]

[1]Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, UK

[2]Meteorological Observatory Hohenpeißenberg, German Weather Service, Germany

[3]Met Office, Exeter, UK

[4]European Centre for Medium-Range Weather Forecasts, Reading, UK

[5]Finnish Meteorological Institute, Helsinki, Finland

[6]Department of Meteorology, Stockholm University, Sweden

[7]National Centre for Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, UK

Correspondence to: Gillian Young (G.Young1@leeds.ac.uk)

Abstract: By synthesising remote-sensing measurements made in the central Arctic into a model-gridded Cloudnet cloud product, we evaluate how well the Met Office Unified Model (UM) and European Centre for Medium-Range Weather Forecasting Integrated Forecasting System (IFS) capture Arctic clouds and their associated interactions with the surface energy balance and the thermodynamic structure of the lower troposphere. This evaluation was conducted using a four-week observation period from the Arctic Ocean 2018 expedition, where the transition from sea ice melting to freezing conditions was measured. Three different cloud schemes were tested within a nested limited area model (LAM) configuration of the UM – two regionally-operational single-moment schemes (UM_RA2M and UM_RA2T), and one novel double-moment scheme (UM_CASIM-100) – while one global simulation was conducted with the IFS, utilising its default cloud scheme (ECMWF_IFS).

Consistent weaknesses were identified across both models, with both the UM and IFS overestimating cloud occurrence below 3 km. This overestimation was also consistent across the three cloud configurations used within the UM framework, with >90% mean cloud occurrence simulated between 0.15 and 1 km in all model simulations. However, the cloud microphysical structure, on average, was modelled reasonably well in each simulation, with the cloud liquid water content (LWC) and ice water content (IWC) comparing well with observations over much of the vertical profile. The key microphysical discrepancy between the models and observations was in the LWC between 1 and 3 km, where most simulations (all except UM_RA2T) overestimated the observed LWC.

Despite this reasonable performance in cloud physical structure, both models failed to adequately capture cloud-free episodes: this consistency in cloud cover likely contributes to the ever-present near-surface temperature bias simulated in every simulation. Both models also consistently exhibited temperature and moisture biases below 3 km, with particularly strong cold biases coinciding with the overabundant modelled cloud layers. These biases are likely due to too much cloud top radiative

33   cooling from these persistent modelled cloud layers and were interestingly consistent across the three UM configurations tested,

34   despite differences in their parameterisations of cloud on a sub-grid-scale. Alarmingly, our findings suggest that these biases

35   in the regional model were inherited from the driving model, thus triggering too much cloud formation within the lower

36   troposphere. Using representative cloud condensation nuclei concentrations in our double-moment UM configuration, while

37   improving cloud microphysical structure, does little to alleviate these biases; therefore, no matter how comprehensive we make

38   the cloud physics in the nested LAM configuration used here, its cloud and thermodynamic structure will continue to be

39   overwhelmingly biased by the meteorological conditions of its driving model.

1   Introduction

41   The Arctic is warming at more than twice the global average rate (**Serreze and Barry, 2011**; **Cohen et al., 2014**), with recent

42   evidence suggesting the rate of warming could be up to three times the global average (**AMAP 2021**). Coupled general

43   circulation models (GCMs) fail to agree on the magnitude of recent warming and exhibit large biases in surface temperature

44   and energy balance (**Boeke and Taylor, 2016**) driven largely by model parameter uncertainties on a decadal scale (**Hodson et**

45   **al., 2013**). Biases in such surface properties are also present in atmosphere-only versions of these models with fixed ocean and

46   sea ice boundaries, indicating that there is an important atmospheric source of disparity between models and reality (**Bourassa**

47   **et al., 2013**). Arctic clouds have a net warming effect at the surface (**Boucher et al., 2013**) and are likely a contributing factor

48   to the spread of surface energy balance estimates obtained from these models, with a large spread in cloud fractions, liquid

49   water paths (LWP), and ice water paths (IWP) identified in past phases of the Coupled Model Intercomparison Project (CMIP;

50   **Karlsson and Svensson, 2011**; **Boeke and Taylor, 2016**). Early results from the most recent CMIP indicate that high latitude

51   discrepancies in cloud fraction are still prevalent in recent revisions of these models (**Vignesh et al., 2020**).

52   With accelerating Arctic warming, we need to build suitable numerical models to confidently predict how the atmosphere will

53   change both on short weather prediction and longer climate time scales (**Jung et al., 2016**). Models such as the Met Office

54   Unified Model (UM) and European Centre for Medium-Range Weather Forecasting (ECMWF) Integrated Forecasting System

55   (IFS) are commonly used for assessing future Arctic change; however, recent work has shown that, like other large-scale

56   models, both exhibit surface energy balance discrepancies with comparison to high Arctic observations. In both the UM and

57   the IFS, these biases have largely been attributed to incorrect cloud cover (**Birch et al., 2012**; **Sotiropoulou et al., 2016**;

58   **Tjernström et al., 2021**).

59   Several studies have considered why such large-scale models fail to reproduce observed cloud cover in the high Arctic.

60   Observations have shown that during summer, Arctic clouds experience episodes of extremely low concentrations of cloud

61   condensation nuclei (CCN; < 10 cm^{-3}) approximately 10—30% of the time (**Mauritsen et al., 2011**; **Tjernström et al., 2014**),

62   highlighting that model capability to reproduce cloud free conditions in the Arctic is likely dependent upon representing these

63   low CCN numbers (**Birch et al., 2012**; **Stevens et al., 2017**; **Hines and Bromwich, 2017**). Such conditions are difficult to

64   simulate with large-scale numerical models utilising single-moment microphysics schemes with assumed constant droplet

65   number concentrations, N_d. Both the IFS and the UM make such assumptions in their current global operational configurations:

66   while climatological aerosol concentrations are referenced in the calculations of the first and second indirect effect, droplet

67   number cannot evolve independently of cloud liquid mass.

68   The operational single-moment microphysics scheme within the UM was found to hinder its ability to reproduce tenuous cloud

69   periods during the *Arctic Summer Cloud Ocean Study* (ASCOS); when clouds were modelled, the model produced too thin

70   cloud layers in a boundary layer (BL) that was often too well-mixed and too shallow (**Birch et al., 2012**). The prevalence of

71   too much low-level cloud caused surface energy balance, and hence surface temperature, biases. The new Cloud-Aerosol

72   Interactive Microphysics (CASIM) double-moment scheme in the UM has enabled improvements in its representation of Arctic

73   clouds; **Stevens et al. (2017)** noted it improved the surface net longwave radiation (LW_{net}) in both cloudy and cloud-free

74   conditions. Specifically, inclusion of aerosol processing within CASIM successfully led to cloud dissipation when modelling

75   the CCN-limited clouds observed during the ASCOS campaign, indicating that this explicit description of double-moment

76   microphysics (rather than a simplified cloud physics description) is key to modelling these clouds in the high Arctic.

77   Like the UM, the IFS also failed to capture episodic cloud-free periods observed during ASCOS, leading to similar surface

78   energy biases (**Sotiropoulou et al., 2016**). The updated IFS cloud scheme, used operationally since 2013, has improved its

79   ability to capture mixed-phase Arctic clouds in recent revisions; however, **Sotiropoulou et al. (2016)** reported that the IFS still

80   exhibits a persistent positive near-surface temperature bias, despite the improvement to its representation of these clouds. These

81   Arctic surface biases persist in version Cy45r1 of the model, as shown by **Tjernström et al. (2021)**. Given that reanalysis

82   products created using the ECMWF IFS (e.g., ERA5; **Hersbach et al. 2020**) are widely used, both to produce lateral boundary

83   conditions for process studies with numerical weather prediction (NWP) models and to analyse Arctic atmospheric structure,

84   we must understand the root of these biases and make recommendations for process improvements.

85   Here, we evaluate the performance of recent revisions of both the UM and IFS focusing on their ability to capture clouds and

86   the thermodynamic structure of the BL, highlighting common process relationships between the models which may explain

87   differences from observations. To achieve this, we compare these models with recent high Arctic observations made during the

88   Arctic Ocean 2018 (AO2018; **Vüllers et al., 2021**) expedition, where a suite of remote-sensing instrumentation was active

89   aboard the Swedish icebreaker *Oden* measuring summertime cloud and BL properties in the high Arctic. We use Cloudnet

90   (**Illingworth et al., 2007**) to compare observations with cloud properties simulated by the models, to test the respective

91   components in each model simulation with a focus on evaluating the relative contributions of the following on cloud structure:

92   1.   The choice and use of large-scale cloud schemes at high resolution

93   2.   The cloud microphysics scheme chosen to represent resolved clouds

94   3.   Representative CCN concentrations, and thus droplet number concentrations, as a function of altitude

95   4.   The global model analyses used to produce boundary conditions for high resolution nested configurations

96   By testing these hypotheses with two different atmospheric models, operating on different grid configurations, we assess

97   whether representative CCN concentrations are indeed the key model development still required to suitably capture Arctic

98   clouds, or whether other factors are restricting model performance in the high Arctic.

2 Data and methods

2.1 Arctic Ocean 2018 expedition

During the AO2018 expedition, *Oden* drifted with an ice floe near the North Pole from 14 August to 14 September 2018 (**Fig. 1**). Campaign details, instrumentation, and meteorological measurements from the AO2018 expedition are summarised in **Vüllers et al. (2021)**. Here, we use a subset of the measurements for direct comparison with our model simulations. Key acronyms, abbreviations, and parameters referenced in this study are listed in **Table 1**.

Radiosondes (Vaisala RS92) launched at 0000, 0600, 1200, 1800 UTC, provide in situ thermodynamic profiles with a 0.5°C and 5% manufacturer specified uncertainty associated with temperature and humidity sensors, respectively. The radiosonde data were distributed via the global telecommunications system and assimilated operationally at the Met Office and ECMWF. Remote sensing measurements from a Metek MIRA-35 Doppler cloud radar, a Halo Photonics Streamline Doppler lidar, and an RPG HATPRO microwave radiometer were processed through the Cloudnet algorithm (**Illingworth et al., 2007**) following the data preparation steps of **Achtert et al. (2020)**. A Vaisala PWD22 present weather sensor (PWS) measured visibility, precipitation type, precipitation intensity, and cumulative amount; near surface temperature and relative humidity (RH) were obtained from an aspirated Rotronic HMP101 sensor. Broadband downwelling solar and infrared radiation were measured on board the ship by Eppley PSP and PIR radiometers. 3-hourly albedo estimates from surface images were used to calculate upwelling shortwave radiation (**Vuellers et al., 2021**).

2.2 Cloudnet

Cloudnet is used to directly compare between our measured and modelled cloud properties (**Illingworth et al., 2007**). Cloudnet ingests Doppler cloud radar and lidar, ceilometer, microwave radiometer, and radiosonde data to derive cloud fractions and cloud water contents on a chosen model grid. A comprehensive description of the algorithm is beyond the scope of this paper and is provided by **Illingworth et al., 2007 (and references therein),** but essentially the algorithm first homogenizes observational data to a common time resolution of 30 s and interpolates data to the radar height grid. Radar reflectivity (Z_e) and lidar backscatter (β) profiles are used to determine cloud boundaries. Cloudnet takes advantage of the lidar's sensitivity to small particles, such as cloud droplets and aerosol, and the radar's sensitivity to large particles, such as ice particles, rain, and drizzle. Cloud phase is determined using Z_e, β, and thermodynamic information from the radiosondes. Cloud ice water content (IWC) is derived using Z_e and temperature (**Illingworth et al., 2007**), while liquid water content (LWC) is derived by partitioning LWP measured by the radiometer to the identified liquid cloud layers from the lidar. Additionally, an adiabatic LWC is calculated from temperature and humidity profiles and the identified cloud top and base height from radar and lidar measurements.

Cloudnet has already been utilised to study Arctic cloud properties using measurements made aboard *Oden* both during this campaign (**Vüllers et al., 2021**) and during the *Arctic Clouds in Summer Experiment* in 2014 (**Achtert et al., 2020**). Potential errors associated with the Cloudnet procedure are described in **Achtert et al. (2020)**. One particular limitation relevant to this study is the minimum detection height of 156 m (lowest radar range gate). Low level clouds/fog below this height are hence missed by Cloudnet (**Vüllers et al., 2021**) and not included in model comparisons. This limitation also results in problems with

Atmospheric
Chemistry
and Physics
Discussions

Open Access

EGU

133   the LWC derived from radiometer measurements; therefore, we use the calculated LWC under adiabatic assumption in this

134   study (for further details see **Appendix A**)

135   For comparisons with models, cloud fraction by volume (C_V), adiabatic LWC, and cloud IWC from observations are averaged

136   to a reference model grid; here, we use the UM grid, but we could have equally chosen that of the IFS (**Fig. S1**). Cloud properties

137   are calculated using measurement profiles alongside model wind speed and grid-box size, where changes in cloud properties

138   over time are assumed to be driven primarily by advection and not microphysical changes (**Illingworth et al., 2007**). This

139   procedure is applied for C_V, LWC, and IWC, with C_V defined as the fraction of pixels in a 2D slice which are categorised as

140   liquid, supercooled liquid, or ice (**Illingworth et al., 2007**).

141   **2.3**   **Models**

142   A summary of each model simulation is included in **Table 2** and detailed in the following sections. 36-hour forecasts were

143   performed with each model, initialised each day at 1200 UTC with the first 12 hours of spin up discarded, thereby producing

144   daily forecast products (0000 UTC – 0000 UTC) with hourly diagnostics for analysis. This is common practice for such

145   forecasts to ensure discrepancies due to spin up are avoided, while maintaining meteorology close to reality; however, as noted

146   by **Tjernström et al. (2021)**, model error growth is often a function of forecast time and thus the findings of this study may be

147   related to the time window chosen for each model forecast.

148   Column diagnostics from the grid cell closest to the position of *Oden* were extracted from the model domain, updated hourly

149   to account for the ship's drifting position. These variables (e.g., temperature, humidity, cloud fraction, condensate variables,

150   wind versus time) were then used for comparisons with alike variables constructed using Cloudnet (**Illingworth et al., 2007**)

151   with measured data (see **Sect. 2.1**).

152   **2.3.1**   **Integrated Forecasting System (IFS; ECMWF_IFS)**

153   Cycle 46r1 (Cy46r1) of the IFS (used operationally from Jun 2019 to Jun 2020) was used to create global meteorological

154   forecasts. The IFS uses a spectral formulation with a wave-number cut off corresponding to a horizontal grid size of

155   approximately 9 km (**Fig. 1b**). It has 137 levels in the vertical up to 80 km, the lowest at ≈10 m with 8 levels below ≈200 m

156   and 20 below 1 km. IFS forecasts were initialised from ECMWF operational analyses. Operational forecasts produced at the

157   time of the campaign (with Cy45r1) were recently evaluated on a 3-day lead time from a statistical viewpoint for this expedition

158   (**Tjernström et al., 2021**); in contrast, lead-time averaged verification was conducted in this study using a 1-1 comparison of

159   a concatenated timeseries of forecast values (T+13—T+36) with hourly observations.

160   Cloud properties are parameterised following **Forbes and Ahlgrimm (2014)**. This cloud scheme was implemented in Cy36r4

161   and has been previously evaluated for Arctic clouds by **Sotiropoulou et al. (2016)** using Cy40r1. Five independent prognostic

162   cloud variables are included (grid box fractional cloud cover, and specific water contents for liquid, rain, ice, and snow).

163   Heterogeneous primary ice formation is diagnosed following **Meyers et al. (1992)**, with a mixed-phase cut-off of –23 °C.

164   Liquid cloud formation occurs when the average relative humidity within a grid box exceeds a critical threshold, RH_{crit},

165   representing sub-grid-scale variability of moisture. This threshold is 80% in the free troposphere, increasing towards the surface

166 in the boundary layer (**Tiedtke, 1993**). Once formed, cloud liquid mass is distributed across a fixed cloud droplet number
167 concentration, N_d, of 50 cm^{-3} over the ocean (and 300 cm^{-3} over land) to act as a threshold for autoconversion from liquid to
168 rain. For interactions with the radiation scheme, the IFS follows **Martin et al. (1994)** for estimating droplet number, using the
169 prognostic specific liquid water content and a prescribed CCN profile. CCN concentrations are calculated as a function of the
170 near-surface wind speed but decreases with altitude to represent the vertical distribution of aerosol within and above the BL.
171 Further details regarding the cloud scheme can be found in the ECMWF documentation (**IFS Documentation – Chapter 7:**
172 **Clouds and large-scale precipitation, https://www.ecmwf.int/node/19308**).

173 The IFS is coupled to a 0.25° resolution dynamic sea-ice model (Louvain-la-Neuve Sea Ice Model, LIM2) which provides sea
174 ice fractions to the IFS and the surface flux tiling scheme (**Buizza et al., 2017; Keeley and Mogensen, 2018**). The surface
175 energy balance over the sea ice fraction is, however, calculated separately from LIM2 using an albedo parameterisation
176 following **Ebert and Curry (1993)** with fixed monthly climatology values interpolated to the actual time, and a heat flux
177 through the ice calculated using a constant sea-ice thickness of 1.5 m.

178 **2.3.2   Unified Model (UM)**

179 The UM was operated as a high-resolution LAM with a 1.5 km × 1.5 km grid (grid is shown in **Fig. 1**). A rotated pole
180 configuration provided approximately equal spacing between grid points towards 90 °N. The LAM contained 500 × 500 grid
181 boxes, spanning from 83.25 °N to 90 °N centred on the 30 °E meridian. In the vertical, there were 70 vertical levels up to 40
182 km, with 24 levels within the lowest 2 km of the domain (**Grosvenor et al., 2017**). Lateral boundary conditions were generated
183 hourly from UM global model 36-hour forecasts at N768 resolution (corresponding to approximately 17 km at 90 °N with the
184 rotated pole) using the Global Atmosphere 6.1 configuration (**Walters et al., 2017**; **Table 2**). Three configurations of the UM
185 LAM were tested for the main body of this study, each using different combinations of cloud microphysics and large-scale
186 cloud schemes. Details on the pertinent microphysical processes represented in each simulation are listed in **Table 3**.

187 **2.3.2.1   Regional Atmosphere model configurations (UM_RA2M and UM_RA2T)**

188 Version 2 of the Regional Atmosphere model within the UM framework has two standard configurations: the mid-latitude
189 configuration (UM_RA2M) and the tropical configuration (UM_RA2T). Both are used operationally in their respective
190 geographical regions. The key difference between these configurations can be found in their turbulent mixing processes:
191 UM_RA2M employs weak turbulent mixing to encourage heterogeneity in model fields to facilitate the triggering of small
192 convective showers; however, while this weak mixing works well to reproduce conditions often experienced in the mid-
193 latitudes, it triggers convection too early in the tropics. Therefore, these two standard Regional Atmosphere configurations
194 were designed separately to account for these subtle differences in convection initiation on km-scales (**Bush et al., 2020**).

195 Neither configuration has been previously evaluated for use in the Arctic. Note that both UM_RA2M and UM_RA2T use the
196 default Regional Atmosphere surface albedo thresholds, giving a 50% albedo at an ice surface temperature of 0 °C and
197 increasing to 80% at –10 °C. **Gilbert et al. (2020)** tested both configurations for polar cloud modelling over the Antarctic
198 Peninsula, finding that the mid-latitude scheme performs better than the tropical configuration for capturing polar cloud liquid

199 water properties and associated radiative interactions (with the surface albedo modelled to within 2% of observed values),

200 whereas the tropical scheme enabled a too-efficient ice phase (and associated liquid depletion).

201 Both UM_RA2M and UM_RA2T include the **Wilson and Ballard (1999)** description of large-scale precipitation to simulate

202 resolved cloud microphysics. This microphysics scheme describes prognostic liquid and ice mixing ratios (q_{liq} and q_{ice},

203 respectively), with an assumed fixed N_d profile calculated from an aerosol climatology and tapered to 50 cm^{-3} towards the

204 surface (between 150 m and 50 m). A single ice species (encapsulating pristine crystals, aggregates, and snow particles) is

205 represented, with an assumed particle size distribution based on **Field et al. (2007)**.

206 UM_RA2M uses the **Smith (1990)** large-scale cloud scheme to parameterise sub-grid-scale fluctuations in humidity and cloud,

207 designed to ensure coarse grid GCMs do not have entirely cloudy grid boxes. It diagnoses cloud fraction and condensate

208 variables for input to the microphysics scheme, referencing a prescribed RH_{crit} profile (based on a symmetric triangular

209 probability density function of sub-grid-scale variability in temperature and moisture) to permit condensation below 100 %

210 humidity (**Wilson et al., 2008**). Condensation cannot occur within a grid box until the grid box mean RH exceeds a prescribed

211 altitude-dependent RH_{crit} (described in **Table 3**).

212 In UM_RA2T, the prognostic cloud fraction and prognostic condensate (PC2; **Wilson et al., 2008**) large-scale cloud scheme

213 is used, designed to address the over-sensitive diagnostic links between cloud fraction and cloud condensate in **Smith (1990)**.

214 Total, liquid, and ice cloud fractions are included as prognostic variables in PC2, allowing cloud fractions and condensate to

215 vary through other interactions (such as BL processes and cloud microphysics) and not simply diagnosed from temperature and

216 humidity as in **Smith (1990)** (**Wilson et al., 2008**). PC2 prognostic variables are advected by the wind and continually updated

217 following incremented sources and sinks in the model, with the additional inclusion of sub-grid-scale turbulent production of

218 liquid in mixed-phase cloud from an analytical model of sub-grid-scale moisture variability (**Furtado et al., 2016**). Differences

219 between the methods of representing cloud fraction in the PC2 and Smith schemes are detailed in the **Supporting Information**.

220 **2.3.2.2 Regional Atmosphere with Cloud-Aerosol Interactive Microphysics scheme (UM_CASIM-100)**

221 UM_CASIM-100 uses the CASIM scheme (detailed by **Hill et al., 2015**) coupled with the **Smith (1990)** large-scale cloud

222 scheme (as in **Grosvenor et al., 2017**). **Stevens et al. (2017)** previously tested the CASIM scheme within the UM nesting suite

223 in an Arctic cloud case study, showing that it performed well in capturing cloud dissipation; however, the authors did not

224 include sub-grid-scale contributions from **Smith (1990)** in that study.

225 CASIM utilises prescribed lognormal aerosol distributions to provide a double-moment representation of cloud particle

226 processes and is the only double-moment setup included in this study. Particle size distributions of five hydrometeors (liquid

227 droplets, ice, snow, graupel, and rain) are each described by a gamma distribution, with prognostic mass mixing ratios and

228 number concentrations. Ice number concentrations are diagnosed via a temperature-number concentration parameterisation

229 (**Cooper, 1986**), but require liquid to be present before ice can form; a relationship thought to be important in Arctic mixed-

230 phase clouds (e.g., **de Boer et al., 2011; Young et al., 2017**). Droplet activation follows **Abdul-Razzak and Ghan (2000)**,

231 referencing a fixed soluble accumulation mode aerosol number concentration profile of 100 cm^{-3}. This profile was

232   approximated based on aerosol concentration profiles previously measured during summertime in the central Arctic

233   (**Kupiszewski et al., 2013**).

234   CASIM offers user flexibility regarding aerosol processing, as described by **Miltenberger et al., (2018)**. Here we do not impose

235   wet-scavenging processes, likely important for capturing cloud-free conditions, for consistency with the simpler single-moment

236   liquid microphysics schemes used in the other simulations; however, use of this option will be explored in future work.

237   For our CASIM simulation, we adapt the warm ice temperature albedo of the LAM to 72% (at 0 °C), with 80% albedo achieved

238   at –2 °C, to match the parameterisation limits currently used in the *Joint UK Land Environment Simulator* surface scheme of

239   the Global Atmosphere 6.0 global model (under the assumption that snow is present on the sea ice surface). For the drift period,

240   we know that snow was indeed present on the surface from first-hand knowledge and surface imagery; therefore, we use this

241   simulation to test the effect of such an increased albedo at warmer surface ice temperatures on the modelled surface energy

242   balance. An example simulation utilising the CASIM scheme with the default Regional Atmosphere albedo settings used in

243   UM_RA2M and UM_RA2T, to demonstrate the radiative impact of CASIM alone, is described in the **Supporting Information**

244   **(Sect. S2).**

245   **2.4**   **Comparison methodology and compared parameters**

246   C_V, q_{liq}, and q_{ice} from each model simulation were ingested by Cloudnet to calculate LWC and IWC. Within these calculations,

247   Cloudnet filters model data for values outside the range observable by the instrumentation used; for example, q_{ice} data are

248   filtered for values which would be beyond the observable range of the radar.

249   We use an additional metric alongside C_V based on total condensate for comparisons between our measured and modelled

250   clouds; a total water content (TWC) mask where the grid-box is considered cloudy; this mask is set to 1, when $TWC \geq 1 \times 10^{-6}$

251   kg m^{-3} below 1 km, and $TWC \geq 1 \times 10^{-7}$ kg m^{-3} above 4 km, with vertical interpolation in between (following **Tjernström et**

252   **al., 2021; Fig. S3**). While this mask will not capture fractional cloud at cloud boundaries, averages of this mask are directly

253   comparable between the observations and models. It acts as a comparison metric based solely on cloud water contents, which

254   are prognostic in every simulation, and does not depend on a specific definition of e.g., cloud fraction.

255   In addition to a full overview of model performance over the drift, we further split our data into sub periods to aid our

256   interpretation of the comparisons between the measurements and models. The sea ice melt/freeze transition was captured by

257   the measurements; **Vüllers et al. (2021)** identified the sea ice freeze onset date as 28 Aug and defined sub-periods throughout

258   the drift based on consistent meteorology (see **Fig. 2g**). We concentrate on the sea ice melt and freeze periods separately and

259   on shorter episodes within these periods; one during the sea ice melt (14—18 Aug) and one during the freeze (4—8 Sept).

260   **3**   **Results**

261   **3.1**   **Surface radiation**

262   **Figure 2** shows measured and modelled time series of net surface shortwave (SW_{net}) radiation, LW_{net}, and the combined

263   surface net radiation (R_{net}) during the AO2018 drift period. All radiative quantities are defined as positive downwards.

264 All models overestimate SW_{net} (**Fig. 2a**) with respect to measurements, with ECMWF_IFS and UM_CASIM-100 in better

265 agreement with observations than UM_RA2M and UM_RA2T. From **Fig 2d**, all simulations fail to capture strong longwave

266 net emission likely related to cloud-free episodes (e.g., 20—21 Aug) and sporadically predict such cloud-free conditions (and

267 net longwave emission) when clouds were observed (e.g., 2 Sep).

268 Considering the melt and freeze periods separately, the measured R_{net} is often negative after 28 Aug (**Fig. 2g**) driven by LW_{net},

269 while SW_{net} decreases with the declining solar elevation angle (**Fig. 2a**). In contrast, the models' net radiation is not typically

270 negative until after 8—9 Sep, excluding a short negative period at 2 Sep driven by the lack of modelled cloud (as suggested by

271 strong net longwave emission; **Fig. 2d**). This delay would likely affect the freeze onset if the models were fully coupled to a

272 sea ice model; as such, this feedback may be active within the (simple) coupled atmosphere-sea ice system of the IFS.

273 Probability Density Functions (PDFs) of these data, split between melt and freeze periods (**Fig. 2b—c, e—f, h—i**), reveal some

274 clear distinctions in model capability. SW_{net} PDFs vary substantially between the models during the melt period (**Fig. 2b**); no

275 simulation captures the observation distribution well. Observed SW_{net} from the ship has a median of +18.2 W m^{-2}, with each

276 simulation producing medians at greater values (UM_CASIM-100 = +18.7 W m^{-2}; ECMWF_IFS = +21.6 W m^{-2}; UM_RA2M

277 = +40.5 W m^{-2}; and UM_RA2T = +41.9 W m^{-2}). While the medians for UM_CASIM-100 and ECMWF_IFS are in good

278 agreement with observations, both exhibit a too-narrow distribution. These too-narrow distributions – which also all lack a very

279 high positive tail – suggest that the modelled cloud cover is too consistent, likely related to the lack of cloud-free episodes

280 indicated by the LW_{net} data (**Fig. 2d**). Median SW_{net} of both the UM_RA2T and UM_RA2M PDFs is much too high, with

281 non-negligible occurrences > +50 W m^{-2}. The improvement of UM_CASIM-100 over UM_RA2T and UM_RA2M indicates

282 that both the surface albedo used by default in the Regional Atmosphere configurations is too low and the updated cloud physics

283 description of CASIM improves the modelled cloud-radiation interactions. A trial simulation utilising the cloud physics setup

284 of UM_CASIM-100 alongside the default Regional Atmosphere surface albedo parameterisation inputs (as used in UM_RA2M

285 and UM_RA2T) shows that the double-moment cloud physics representation alone does improve radiative properties with

286 comparison to the standard configurations (see **Supporting Information**); however, the combination of improved cloud-

287 radiation interactions and an updated surface albedo (as shown here in UM_CASIM-100) provides the best agreement between

288 the UM and our observations.

289 During the freeze period, measurement estimates of SW_{net} peak at +7.9 W m^{-2}, while ECMWF_IFS, UM_CASIM-100,

290 UM_RA2M, and UM_RA2T have maxima at +10.0, +10.4, +25.0, and +26.6 W m^{-2} respectively (**Fig. 2c**). The peak modelled

291 SW_{net} remains too high in all simulations but, in contrast to the melt period, all PDFs are now too broad. ECMWF_IFS and

292 UM_CASIM-100 perform best with comparison to observations (both with a positive bias of less than +3 W m^{-2} at their peaks).

293 However, both UM_RA2M and UM_RA2T have a broad bimodal structure, with the secondary peak in better agreement with

294 the observations than their maxima. Both UM_RA2T and UM_RA2M are largely in better agreement with observations during

295 the freeze period than during the melt; this improved agreement is likely due to either a better representation of incoming

296 shortwave radiation or the surface albedo; the surface temperatures decreases through the transition to sea ice freezing

297 conditions, and **Fig. S4** indeed shows that the albedo modelled during the freeze for UM_RA2M and UM_RA2T is in better

298 agreement with observational estimates than that modelled during the melt period.

Atmospheric
Chemistry
and Physics
Discussions

299   During the melt period, LW_{net} aligns well between the measurements and models; however, all simulations produce a narrower

300   PDF than the observations and largely miss the tail < —20 W m^{-2} (**Fig. 2e**) resulting from observed cloud-free episodes at 15—

301   16 Aug, 20 Aug, 22 Aug, and 26 Aug (**Fig. 2d**). Despite this, each simulation performs well in replicating the median of the

302   PDF, with a maximum model-observation difference of —1.9 W m^{-2} (UM_RA2T). As with SW_{net}, model-observation

303   agreement generally improves during the freeze period, with UM_RA2M, UM_RA2T, and ECMWF_IFS producing PDFs

304   closely matching the observations, with median values at —9.4, —11.5, and —6.8 W m^{-2} respectively compared with an

305   observation peak of —6.5 W m^{-2} from the ship estimates. Each of these cases also reproduces the negative distribution tail

306   missed by all simulations during the melt (**Fig. 2f**). UM_CASIM-100 displays a narrower distribution with fewer negative

307   values, yet still performs equally well in reproducing the median of the LW_{net} PDF (with a bias of —5.5 W m^{-2}). With the

308   exception of the too-narrow UM_CASIM-100 PDF, this improved agreement in LW_{net} indicates that cloud cover is indeed

309   captured better by the models during the freeze, and remaining discrepancies in the SW_{net} comparisons may indeed be related

310   more so to cloud microphysical structure or surface properties.

311   To investigate this relationship in more detail, we split our radiation data into the periods of consistent meteorology indicated

312   on **Fig. 2g.** In agreement with **Fig. 2**, model SW_{net} and downwelling shortwave radiation ($SW_\downarrow$) biases are at their greatest

313   during period 3 (**Table 4**). All simulations similarly exhibit their greatest LW_{net} biases during period 3 (**Table 4**); less cloud

314   cover was observed during this period with relation to other periods during the drift (**Vüllers et al., 2021**). Both UM_RA2T

315   and UM_RA2M perform best in terms of $SW_\downarrow$ during period 3; however, both perform particularly poorly in SW_{net}, indicating

316   that their surface albedo is not representative of observations. During period 4, UM_CASIM-100 performs very well in terms

317   of these $SW_\downarrow$ biases, suggesting the modelled cloud structure was in good agreement with observations. However, ECMWF_IFS

318   exhibits the smallest $SW_\downarrow$ biases of the four simulations during periods 5 and 6; periods when the sea ice was beginning to

319   freeze.

320   Each of these simulations highlight that small $SW_\downarrow$ biases do not necessarily produce similarly small SW_{net} biases, as both the

321   modelled cloud properties and surface albedo need to be representative to remedy the SW_{net} discrepancies. In UM_RA2T and

322   UM_RA2M, the surface albedo is poorly captured, as indicated by the consistently high SW_{net} biases; however, ECMWF_IFS

323   and UM_CASIM-100 perform better in terms of surface albedo, with UM_CASIM-100 performing the best with the smallest

324   SW_{net} biases across the four sub-periods considered. Further discussion of the surface albedo comparison is included in the

325   **Supporting Information**.

326   LW_{net} biases do not exceed +5.5 W m^{-2} over periods 4—6; however, biases are greater (up to +16.3 W m^{-2} during period 3;

327   **Table 4**) due to the models' inability to reproduce cloud-free conditions. This relationship with cloud cover influences the

328   surface downwelling longwave ($LW_\downarrow$) biases: with the exception of the standard UM configurations during period 5, all $LW\downarrow$

329   biases are positive (**Table 4**).

330   Combining these radiative components, we find that R_{net} is overestimated by all simulations during the melt (with

331   UM_CASIM-100 and ECMWF_IFS performing better than UM_RA2M and UM_RA2T; **Fig. 2h**), largely driven by too much

332   surface SW_{net} when cloud is present in reality, thus indicating that the model surface albedo is too low and thus does not reflect

333   enough $SW_\downarrow$. On the other hand, there are also non-negligible occurrences of too much modelled cloud when the conditions

334   should be cloud-free, driving strong LW_{net} biases at these times. While agreement with observations largely improves during

335   the freeze period, these discrepancies still exist in the SW_{net} data. While the SW_{net} biases may be strongly influenced by errors

336   in the surface albedo, and thus beyond the scope of this study, the role of cloud structure in $SW_\downarrow$ biases and the LW_{net} emission

337   episodes missed by each simulation are driven by the description of cloud: in the following sections, we investigate the cloud

338   macro- and microphysical structure and surface properties to explain these radiative differences.

339   **3.2   Cloud properties**

340   To evaluate model performance, we use two metrics for cloud occurrence: the model diagnosed cloud fraction, C_V, and the

341   cloud occurrence inferred from cloud water contents, the TWC cloud mask. **Figure 3** shows TWC and cloud fraction, C_V,

342   calculated from observations using Cloudnet and output by models. TWC comparisons indicate that each simulation captures

343   the observed cloud aloft, except for UM_CASIM-100 between 4 and 10 Sep. Below 3 km, observed TWC is generally lower

344   in magnitude than the model simulations.

345   In contrast, all simulations except UM_RA2T fail to reproduce the observed C_V aloft. Low-altitude (below 2 km) cloud cover

346   appears to be captured comparatively better across all simulations. Cloud height simulated by ECMWF_IFS is in reasonable

347   agreement with the observations; however, there are notable periods where the persistence of clouds aloft is not reproduced.

348   For example, the altitude and timing of onset of the (likely precipitating) high clouds at 3—4 Sep is initially captured, but the

349   clouds are not sustained. Cloud layers aloft appear more tenuous also in UM_RA2M and UM_CASIM-100 than in the

350   observations: there are few cases of cloud fractions > 0.5 at altitudes above 3 km.

351   **Figure 4a** shows mean profiles of C_V over the drift period. Only periods where we have measurement data are included in these

352   profiles for fair comparison. Note that cloud fraction below 0.15 km is not evaluated against observations here due to low-

353   altitude measurement limit of the cloud radar. Supporting qualitative interpretation of **Fig. 3**, model-observation agreement of

354   C_V is best at low altitude (below 1 km); however, all simulations produce too much very low (between 0.15 and 0.5 km) cloud.

355   Modelled near-surface C_V (between 0.15 and 0.5 km) is up to 16% too high (UM_RA2T). However, we can speculate that the

356   frequent fog episodes reported during the ice drift **(Vüllers et al., 2021)** may be somewhat captured by the models, as indicated

357   by mean values of C_V below 0.15 km of 82%, 72%, 53%, and 39% respectively for UM_RA2T, UM_RA2M, UM_CASIM-

358   100, and ECMWF_IFS. All simulations except UM_RA2T perform poorly aloft: ECMWF_IFS, UM_RA2M, and

359   UM_CASIM-100 strongly underestimate C_V between 1 km and 8 km, with UM_CASIM-100 and UM_RA2M reproducing less

360   than 20% of the observed C_V at 4.5 km. Only the UM_RA2T C_V profile agrees well at altitude, with particularly good agreement

361   between 0.5 and 2 km. In fact, C_V between 2 km and 5.5 km agrees best with observations out of the four simulations considered.

362   **Figure 3** highlights that the observations, UM_RA2T, and (to an extent) ECMWF_IFS have a C_V field scaling largely as either

363   0 or 1, whereas UM_RA2M and UM_CASIM-100 are more likely to have a fractional cloud cover aloft, thus producing a poor

364   comparison with our observations (**Fig. 4a**). Despite this, qualitative model-observation comparisons of TWC indicate that the

365   models are performing well. Further discussion of these differences is included in the **Supporting Information**. In summary,

366   the Cloudnet calculation of C_V from observations is not directly equivalent to our model cloud fractions and such comparisons,

367   in isolation, should be approached with caution in the Arctic. To bypass this issue, we also use a cloud mask built from TWC

368   data to aid interpretation of our results. The observed TWC cloud mask (**Fig. 4b**) differs from the mean C_V profile, with a subtle

369   bimodal structure peaking at approximately 0.5 and 4.5 km (with a minimum around 2 km).

370   All simulations overestimate cloud occurrence below 2.5 km (**Fig. 4b**), in contrast to the underestimation between 1 and 2.5

371   km shown in the C_V data (**Fig. 4a**). Mean observed cloud occurrence only reaches 75% between 0.15 km (lowest radar range

372   gate) and 0.5 km, while UM_RA2M and UM_CASIM-100 have more than 98% cloud occurrence at 0.2 km. UM_RA2T

373   performs slightly better, peaking to only 92% at 0.2 km; however, the improvement is not as significant between UM_RA2T

374   and UM_RA2M/UM_CASIM-100 as is suggested by the mean C_V profiles (**Fig. 4a**). ECMWF_IFS peaks at a slightly higher

375   altitude, overestimating cloud occurrence by 33% at approximately 0.5 km (**Fig. 4b**).

376   Above 2 km, ECMWF_IFS, UM_RA2M, and UM_RA2T perform similarly; the greatest difference aloft occurs at 4.5 km,

377   where there is a minor peak in the mean observed cloud occurrence (up to 41%; **Fig. 4b**). ECMWF_IFS produces only 28%

378   cloud cover at this altitude. UM_CASIM-100 cloud occurrence monotonically decreases with altitude above 3.5 km, producing

379   only 20% cloud cover at 4.5 km, in agreement with the qualitative findings of **Fig. 3**. Therefore, with the exception of

380   UM_CASIM-100, the TWC cloud masks indicate that modelled cloud occurrence aloft is, in fact, in reasonable agreement with

381   observations, in contrast to the trends indicated by the C_V data (**Fig. 4a**). These data suggest that the C_V comparisons may be

382   misleading if used in isolation, likely due to the different methods for representing cloud fractions and associated sub-grid-

383   scale variability in models (see **Supporting Information**).

384   Averaged in-cloud water content profiles are shown in **Fig. 4c—d**. Adiabatic LWC calculated from observations with Cloudnet

385   is shown in **Fig. 4c**. This adiabatic assumption was used in place of the HATPRO LWP due to the data quality issues introduced

386   to the latter because of the frequent occurrence of fog at altitudes below the lowest radar range gate (0.15 km; discussed further

387   in **Appendix A**).

388   The adiabatic LWC peaks between 0.5 and 1 km then decreases steadily with altitude between 1 and 3 km. All simulations

389   overestimate in-cloud LWC between 1 and 3 km; however, below 1 km, each simulation (except UM_RA2T) performs

390   reasonably well. At 0.5 km, UM_RA2T underestimates by 47 %, while UM_CASIM-100 overestimates by just 10 % and

391   UM_RA2M and ECMWF_IFS are in reasonable agreement with observations. UM_RA2T and UM_RA2M have bimodal

392   distributions, with peaks below 0.5 km and around 2 km, perhaps linked to their common use of the **Wilson and Ballard (1999)**

393   microphysics scheme. The increase in LWC towards the surface in UM_RA2M is suggestive of fog, and UM_RA2M is the

394   only simulation to display this vertical structure. The mean LWC calculated for ECMWF_IFS is consistent with altitude

395   between 0.5 and 2 km; however, there is more variability at 2 km than at lower altitudes, indicating that this may be a more

396   dominant liquid cloud layer at some time periods. Only UM_CASIM-100 displays a similar shape to the observations, yet its

397   LWC is often greater than the observed LWC at all altitudes above 1 km.

398   All simulations agree with the Cloudnet-calculated IWC above 4 km (**Fig. 4d**); in fact, UM_RA2M performs particularly well

399   across the entire vertical profile. ECMWF_IFS and UM_CASIM-100 also agree well for most of the profile apart from slight

400   overestimations below 1.5 km (though still within one standard deviation of the observed mean). UM_RA2T overestimates

401   below 4 km, producing almost seven times the observed IWC (0.019 g m^{-3} versus 0.003 g m^{-3}) at 0.5 km. Shaded standard

402   deviations also indicate that UM_RA2T is also more variable than both the three other simulations and the measurements,

403   consistent with previous studies showing its ice phase is more active than UM_RA2M in polar mixed-phase clouds (**Gilbert**

404   **et al., 2020).**

405   Column-integrated metrics and surface measurements provide an additional perspective for evaluating model performance with

406   regards to clouds. Measured LWP and precipitation fluxes are shown alongside corresponding model diagnostics in **Fig. 5**.

407   Cloudnet-filtered LWP is included in **Fig. 5a,b** for comparison; these data are HATPRO measurements filtered by Cloudnet

408   for bad points (e.g., strong precipitation events). ECMWF_IFS, UM_RA2M, and UM_CASIM-100 produce $LWPs$ in

409   reasonable agreement with measurements throughout the full drift period, with the PDFs of **Fig. 5b** indicating that these $LWPs$

410   are overestimated slightly with respect to the measurements/Cloudnet data. UM_RA2M overestimates in some periods, for

411   example the LWP peak during the storm of 12 Sep is 230 g m^{-2} more than measured (**Fig. 5a**). In contrast, UM_RA2T

412   underestimates the LWP overall, with few occurrences of > 200 g m^{-3} (**Fig. 5b**). This underestimation of LWC (**Fig. 4c**) and

413   LWP (**Fig. 5a, b**) by UM_RA2T aligns with its overestimation of IWC below 4 km; with too much ice in mixed-phase cloud,

414   liquid is depleted too efficiently via the Wegener-Bergeron-Findeisen mechanism.

415   Each simulation broadly captures the notable precipitation events measured (**Fig. 5c—d**). UM_CASIM-100 and UM_RA2T

416   reproduce the measured total precipitation flux well and capture the short episodes of increased precipitation at 22 Aug, 3 Sep,

417   and 12 Sep. ECMWF_IFS and UM_RA2M also capture some precipitation events; however, the magnitude of these events is

418   best reproduced by UM_CASIM-100. No simulation reproduces the precipitation intensity measured at 8 Sep. While the key

419   precipitation events are largely captured by the models, with each model producing precipitation as predominantly snow rather

420   than rain, the precipitation rates simulated are low and likely contribute to the lack of cloud-free periods as indicated by the

421   LW_{net} comparisons shown previously (**Fig. 2d, e, f**).

422   These results therefore indicate that the modelled microphysical structure is positively biased in terms of cloud liquid with

423   respect to observations (**Figs. 4c, d—5**). There is a consistent model-observation bias, with all simulations producing too much

424   cloud (**Fig. 4a, b**) below 2.5 km. In ECMWF_IFS, UM_RA2M, and UM_CASIM-100, this cloud contains too much liquid (as

425   indicated by positive biases in LWC and LWP). Only UM_RA2T underestimates the cloud liquid properties due to its active

426   ice phase. **Figure 6** links the radiation, LWP, and C_V biases of our four model simulations with respect to observations. C_V

427   biases are calculated as the model-observation bias below 3 km, where model data below the height of the lowest radar range

428   gate (0.15 km) are excluded from the bias calculation. Here, C_V is used in place of the TWC cloud mask as the latter is calculated

429   from in-cloud LWC and is therefore not strictly independent of LWP. These linear regressions demonstrate that the positive

430   downwelling radiative biases are indeed tied to too much cloud cover within the models, and too much liquid within the

431   modelled clouds. The correlations are weaker for R_{net}, likely due to the additional influence of other factors (e.g., surface

432   albedo) on the net radiative properties.

3.2.1   Influence of CCN concentration

434   Each simulation overestimates cloud occurrence below 2.5 km and struggles to maintain cloud-free conditions, problems

435   previously identified for earlier versions of these models. Both **Sotiropoulou et al. (2016)** and **Birch et al. (2012)** commented

436   on the need for variable, representative cloud nuclei concentrations in the IFS and the UM to enable cloud-free periods to be

437   captured. A fixed accumulation-mode aerosol number and mass concentration profile was used in UM_CASIM-100; however,

438   such consistency with altitude is unlikely to occur in reality. While the concentration chosen was based on previous

439   measurements in the Arctic (**Kupiszewski et al., 2013**), aerosol number concentrations are typically very low and

440   heterogeneous within the BL during the Arctic summer (**Mauritsen et al., 2011; Tjernström et al., 2014**) yet long-range

441   transport provides comparatively greater, more homogeneous concentrations aloft.

442   An additional simulation with the CASIM scheme was tested, using a more representative CCN vertical profile guided by

443   output from the UK Chemistry and Aerosol (UKCA; **Morgenstern et al., 2009**; **O'Connor et al., 2014**) global model. Details

444   on the UKCA model configuration used to obtain these aerosol data are included in the **Supporting Information**. Using

445   representative aerosol profiles as input to the CASIM scheme (with lower CCN concentrations within the lower troposphere

446   and greater concentrations within the free troposphere; denoted UM_CASIM-AeroProf) can affect the $SW_\downarrow$ as expected via the

447   associated influence on N_d and q_{liq} (**Fig. 7**). Low altitude (below 1 km) clouds have a significantly lower N_d , < 25 cm^{-3}, in

448   UM_CASIM-AeroProf than in UM_CASIM-100. This low N_d is expected from periodic episodes of low CCN in the Arctic

449   BL (**Mauritsen et al., 2011**); cloud residual concentrations of up to 10 cm^{-3} were measured on board *Oden* during the AO2018

450   expedition (**Baccarnini et al., 2021**). However, despite the differences in N_d between these two CASIM simulations, q_{liq} does

451   not differ much as the simulated clouds are not heavily precipitating (and thus cloud lifetime is largely unaffected).). This

452   similarity is also displayed in the diagnosed cloud fractions, related to the comparatively unaffected q_{liq}. Despite the

453   consistency in cloud fractions and q_{liq}, the cloud albedo is subtly lowered (as fewer CCN are available) in UM_CASIM-

454   AeroProf, as shown by the $SW_\downarrow$ comparisons in **Fig. 7a—b**.

455   **3.3   Thermodynamic structure**

456   Differences between modelled and observed cloud properties are likely related to the thermodynamic structure of the

457   atmosphere and how well this is modelled. **Figure 8** shows temperature (T) and water vapour specific humidity (q) from

458   radiosondes and anomalies of each simulation with respect to these measurements. The altitude of the main inversion base

459   identified from the radiosondes is shown in black.

460   Each simulation is typically too cold with respect to observations at altitudes just above the main inversion (*left column*; **Fig.**

461   **8**): this anomaly is a consistent feature throughout the drift period and across models, however it is most prominent at the

462   beginning of the drift. These trends indicate that the altitude of the modelled temperature inversion capping the BL is too high,

463   likely driven by too much BL mixing and the associated too-deep cloud layers modelled in each simulation **(Fig. 4b).** Below

464   the observed inversion, the simulations are typically warmer than measured; for example, at 18 Aug all UM simulations have

465   a particularly strong bias (> 3 K) below the observed main inversion, with ECMWF_IFS exhibiting a similar, but smaller, bias.

466   Above approximately 3 km the T biases are typically smaller in magnitude and variable in sign. All UM simulations display

467   similar differences with respect to the radiosonde measurements; for example, each UM simulation exhibits a strong T bias up

468   to 4.4 K at 6.5 km during 9 Sep.

469   q biases are typically small throughout much of the atmospheric column (*right column*; **Fig. 8**), with some instances of larger

470   biases. These stronger biases are not confined to the lowest 3 km as with the temperature data. Radiosonde humidity data up to

471   22 Aug are variable aloft, and this noise affects the biases calculated over this period. However, a strong moisture bias of >

472   0.90 g m^{-3} is evident between 2 km and 4 km over 20—22 Aug in all UM simulations. Similarly, the dry bias (of up to 1.86 g

473   m^{-3}) across the UM simulations from 2—4 Sep is notable and is also present, to a lesser extent, in ECMWF_IFS (up to 0.82 g

474   m^{-3}).

475   When these data are simplified into median profiles (**Fig. 9**), the characteristic biases exhibited by the models become clearer.

476   **Figure 9(a, c)** shows that the T biases are small above 4 km, with all UM simulations exhibiting a slight warm bias and

477   ECMWF_IFS exhibiting a slight cold bias. Similarly, moisture biases are negligible above 4 km in all simulations (**Fig. 9b, d**).

478   However, below 4 km strong biases emerge.

479   From the surface up to 0.5 km, there is a decreasing positive T bias in all simulations. However, the positive surface T bias is

480   reduced during the freeze period for UM_RA2M and UM_RA2T (from +0.28/+0.31 K to +0.20/+0.14 K, respectively) while

481   it intensifies from +0.52 K (+0.46 K) to +0.90 K (+0.56 K) for ECMWF_IFS (UM_CASIM-100) (**Fig. 9c**).

482   During the melt period, all simulations underestimate the temperature between 1 and 3 km, yet there is a clear bimodal structure

483   evident in each profile with secondary negative peaks at lower altitudes (**Fig. 9a**). ECMWF_IFS remains too cold across a

484   deeper layer than the UM simulations, between 0.4 and 3 km. Both the IFS and the UM exhibit strong (up to –1.54 K) biases

485   at 1.75 km. The negative T bias layers at lower altitudes differ in height between the models, with ECMWF_IFS reaching –

486   0.94 K at 0.85 km while the UM simulations exhibit negligible positive biases at this height. The secondary peak in the UM

487   simulations is in fact lower in altitude, at 0.4—0.5 km. T biases are smaller than during the melt period, reaching up to –1.06

488   K (ECMWF_IFS) between 0.65 km and 1 km, and the negative bias peak at 2 km seen previously is no longer present (**Fig.

489   9c**).

490   Similarly, each simulation exhibits a positive q bias towards the surface. These biases change little between the melt and freeze

491   periods (**Fig. 9b, d**); ECMWF_IFS produces the greatest bias in both periods (+0.31 g m^{-3} during both the melt and freeze),

492   while UM_RA2T produce the lowest (+0.24 g m^{-3} and +0.10 g m^{-3} during the melt and freeze, respectively). ECMWF_IFS is

493   too dry, as well as too cold, between 0.5 and 4 km, while the UM simulations are typically too moist (though variable; **Fig.

494   9b**).

495   There is less variability in the q biases during the freeze period. The UM simulations in particular exhibit only small q biases

496   above 0.5 km (**Fig. 9d**). ECMWF_IFS performs well above 2 km; however, similar to trends identified during the melt, it is

497   again too dry between 0.5 and 2km.

498   **Figure 10** similarly shows the median T and q biases modelled by UM_CASIM-100 and UM_CASIM-AeroProf over the

499   whole AO2018 drift period. Even though the clouds are likely more representative of the high Arctic environment in

500   UM_CASIM-AeroProf than UM_CASIM-100, the thermodynamic biases are largely unchanged from the approximated

501   aerosol input of UM_CASIM-100. We speculate that these biases would perhaps differ more so if the modelled clouds were

502   precipitating strongly in either simulation, thus affecting q_{liq} and cloud lifetime. However, considering each of the UM LAM

503   configurations shown here, there is little variability in their thermodynamic biases despite the differences in their representation

504   of aerosol inputs, cloud microphysics, and large-scale cloud scheme. Interestingly, these biases are shared by the UM global

505   model (UM_GLM, shown in grey; **Fig. 9**) used to generate lateral boundary conditions for each LAM. UM_GLM exhibits

506   similar biases as its high-resolution LAM counterparts, suggesting that these thermodynamic biases are sourced from the
507   driving model itself.

3.3.1   Influence of the UM driving model

509   To investigate how much the large-scale forcing is influencing the UM biases, an additional test was performed over a subset
510   of the drift (31 Aug to 5 Sep) using ERA-Interim to initialise the UM global model (labelled UM_RA2M-ERAI-GLM; **Fig**
511   **11**). This test was designed to evaluate whether the initial conditions of the global driving model, and therefore the associated
512   data assimilation (DA) systems used to derive the operational analyses used for initialisation, are largely responsible for the
513   LAM thermodynamic biases we have found in this study. For this test, we used the UM_RA2M configuration for the LAM,
514   and all global model physics options remained the same as in previous simulations (as described in **Table 2**); the only difference
515   was in the initial conditions of the global model.

516   We find that UM_RA2M-ERAI-GLM exhibits T and q biases following ECMWF_IFS between the surface and 3 km, inheriting
517   the ECMWF_IFS near-surface temperature bias discussed previously (**Fig. 11a**). Over this short time period, the UM
518   simulations do not have this bias. Above 3 km, UM_RA2M-ERAI-GLM follows UM_RA2M and UM_GLM, exhibiting a
519   slight warm bias (0.45 K at 5.5 km) in contrast to the cold bias of ECMWF_IFS (–0.65 K at the same altitude).

520   These results confirm that the UM LAM biases within the lower atmosphere shown in **Figs. 8** and **9** are driven by biases in the
521   large-scale forcing from the global model, which may be a result of the model physics itself or the DA used to produce the
522   operational analyses. Given that the Arctic lacks good observational data coverage, DA systems still rely heavily on their model
523   components when creating the analysis products used for model initialisation; therefore, improved observational data coverage
524   may improve these biases in the DA systems and thus global model initial conditions. In the meantime, a different LAM
525   configuration, with a larger nested domain with lateral boundaries further from the science region of interest may break the
526   relationship between global model and LAM biases shown here.

3.4   Links between cloud properties and thermodynamic biases

528   To better understand how the model thermodynamic biases relate to cloud properties in each simulation, we split our drift
529   period further into four subsections – periods 3 to 6, as illustrated in **Figs. 2 and 8** – to study periods of consistent meteorology.
530   Mean equivalent potential temperature (θ_e) and q profiles measured by radiosondes during these periods are shown in **Fig. 12**.
531   Of the four periods considered, period 3 had cloud-free conditions most often. Periods 5 and 6 were similar; both were cloudy
532   and influenced synoptically by three different low-pressure systems over their duration.

533   Cloud properties and thermodynamic biases during periods 3 and 6 are shown in **Fig. 13** (with similar analysis for periods 4
534   and 5 included as **Fig. S7**). As mentioned previously, mean observed cloud occurrence was lower for period 3 than in any other
535   period during the drift. All simulations overestimate the TWC cloud mask below 2 km, with each UM case producing a bimodal
536   mean profile peaking below 0.5 km and at 1.8 km (**Fig. 13a**). Such bimodality is less clear with ECMWF_IFS; it exhibits a
537   lower layer with cloud top at 1 km and a more prominent than a secondary layer at 1.6 km, although the separation of these
538   layers is not as distinct as in the UM cases. The secondary layer at 1.6 km has a greater LWC than the lower layer, with a peak

539   of 0.14 g m^{-3} (**Fig. 13b**). The bimodal cloud structure is also liquid dominated in the UM simulations, where both peaks reach

540   around 0.1 g m^{-3} (and even exceed this magnitude in the 1.8 km layer), across all three configurations.

541   Considering the corresponding median T biases (**Fig. 13d**), there are clear correlations between negative biases and modelled

542   cloud height, suggesting that cloud top LW cooling is a contributing source of these biases. The lower layer (0.75 km) bias in

543   ECMWF_IFS is particularly striking, reaching –4.45 K, and corresponds with the top of a large fraction of liquid-dominated

544   cloud (**Fig. 13a, b**). The mean LWC modelled at this altitude is over three times greater than was observed, with cloud frequency

545   overestimated by 73%. q biases (**Fig. 13e**) are negligible for ECMWF_IFS between 0.75 and 1 km, yet positive below and

546   above this altitude range. The coinciding overestimation of cloud at these heights indicates that the IFS has simulated too much

547   condensation, driven by the availability of too much moisture. Similarly, all UM simulations exhibit a moist bias between 0.5

548   and 1.6 km, between the modelled cloud layers, and exhibit small dry biases where too much cloud is modelled (e.g., 0.5 km).

549   These results indicate that both models have an excess of water vapour, particularly below 3 km, where negligible/dry biases

550   with comparison to observations are in fact an artefact of too much condensation and resulting cloud cover. This excessive

551   cloud cover, on the other hand, has a negative effect on the temperature bias profile, resulting in strong cold biases.

552   The models are in good agreement with the observed LWC during period 6, with the exception of UM_CASIM-100 which

553   produces double the observed LWC at 0.7 km (**Fig. 13g**). In particular, ECMWF_IFS performs well below 2.5 km in terms of

554   LWC, IWC, and cloud occurrence, with the largest difference in the latter occurring at approximately 0.7 km (100% in

555   ECMWF_IFS in comparison to 79% observed). Consequently, the T biases are smaller during period 6 than period 3 for

556   ECMWF_IFS. However, these T biases are still present (**Fig. 13i**), peaking at –0.96 K at 0.65 km, likely caused by this minor

557   overestimation in cloud cover, albeit with representative microphysics.

558   The magnitude of the T biases for the UM simulations is similar between both periods, likely caused by this model producing

559   up to 100 % cloud cover at low altitude. All UM simulations exhibit stronger T and q biases below 1 km than ECMWF_IFS

560   during period 6 (**Fig. 13i—j**). Strong negative T biases accompany the overestimation of cloud cover in each UM case, and the

561   improved model-observation agreement of LWC by UM_RA2M and UM_RA2T does little to alleviate these biases with

562   comparison to the overestimated LWC of UM_CASIM-100. Simply, there is too much low-altitude (below 1 km) cloud causing

563   too much cloud-top radiative cooling in the model, no matter which representation of cloud microphysics or large-scale cloud

564   is used.

565   However, while the q biases were negligible when ECMWF_IFS exhibited particularly strong T biases during period 3, q

566   biases for the UM become notably negative for the same effect during period 6; this is the largest dry bias simulated over the

567   four periods considered (with periods 4 and 5 included in the **Supporting Information**). The surface q bias for the UM

568   simulations is smaller during period 6 than during period 3, and the tropospheric q bias is positive less often, suggesting the

569   positive moisture bias hypothesised previously (leading to too much condensation and cloud cover) is not ubiquitous in the

570   model. In fact, results shown in **Fig. 13**, and **Fig. S7** for periods 4 and 5, suggest that either the increased synoptic activity or

571   freezing sea ice conditions (or both) of periods 5 and 6 acts to reduce this moist bias in the UM.

572   In summary, both models exhibit strong negative T biases at altitudes coinciding with too much liquid-dominated cloud (e.g.,

573   **Fig. 13a, b, d**), likely caused by the consequent enhancement of cloud-top radiative cooling. q biases improve where cloud is

574 modelled during the melt period (**Fig. 13a, e**), suggesting that the q field was perhaps too moist below 3 km to begin with,

575 leading to too much condensation and excessive cloud cover. However, this hypothesis does not appear to be valid during the

576 freeze nor at altitude, as indicated by the negative q biases above 2.5 km which occur where more cloud was observed than

577 modelled: for example, 2.5 to 4 km during period 3 for all simulations (**Fig. 13a, e**), or 2.5 to 3.5 km for the UM simulations

578 during period 6 (**Fig. 13f, j**). In these instances, our models produce too little cloud as they are too dry to facilitate cloud

579 formation. With underestimated cloud formation, the models are also slightly too warm (approximately 0.3 K) due to the

580 missing radiative cooling occurring at these altitudes in reality.

581 While the model T biases align well with their overestimation of cloud cover, our analysis thus far does not account for the

582 height of the capping inversion. Therefore, incorrect placement of cloud in the models, or a too-deep or too-shallow modelled

583 BL, could be contributing to these biases and thus could affect the interpretation of our results.

584 **Figure 14** shows the strongest temperature inversion base identified from each model simulation and the radiosonde

585 measurements. In each dataset, the strongest inversion below 3 km was identified (following **Vüllers et al., 2021**); if a weaker

586 inversion was modelled at a lower altitude which was closer to the inversion base identified from the radiosonde, the model

587 inversion height was adjusted accordingly. In keeping with previous analysis, radiosonde and IFS data were interpolated to the

588 UM vertical grid for fair comparison; this procedure smooths some high-altitude detail in the radiosonde profiles, such that the

589 strength of some higher-altitude inversions is reduced causing weaker low-altitude inversions to be identified as the primary

590 inversion instead.

591 These results indicate that the strongest (unadjusted) inversion in each simulation is often too high (grey points, **Fig. 14**), and

592 weaker inversions at lower altitude are typically in better agreement with identified inversions from radiosondes. Low inversion

593 bases (below approximately 0.5 km) are consistently overestimated in each simulation, particularly during the melt period (not

594 shown), supporting our previous deduction that the model inversions were often too high. The detection algorithm does fail to

595 capture some inversions, predominantly during the freeze period, and instead underestimates the modelled inversion base

596 during this time window with comparison to measurements (lower right-hand points in each panel).

597 Modelled and observed temperature profiles were scaled using these identified inversions to remove the differences in inversion

598 height from our interpretation of the model biases (**Fig. 15**). When averaged over the full drift, the models are largely biased

599 warm below the inversion and cold above (up to 3 km; **Fig. 15a**), with the exception of UM_CASIM-100 which also exhibits

600 a subtle cold bias just below the inversion. This warm below/cold above signal is more consistent between the models during

601 the melt period (**Fig. 15b**). Above the inversion, ECMWF_IFS exhibits a stronger cold bias than the UM simulations. The

602 shape of the scaled profile is rather consistent between the melt and freeze with ECMWF_IFS; the model is consistently too

603 warm below the inversion, and too cold above, with comparison to radiosonde measurements. However, the UM simulations,

604 particularly UM_RA2M, are partially biased cold below the inversion during the freeze. As previously mentioned, biases during

605 the freeze period must be interpreted with caution as the inversion detection algorithm performed less well during this time

606 window, with several modelled inversions missed. However, these scaled T bias profiles support our previous hypothesis that

607 cloud longwave cooling is producing colder thermodynamic conditions in the models than were observed, irrespective of the

608 differences between modelled and observed inversion heights. Similarly, the warm surface bias indicated previously can be
609 interpreted to span most of the lower troposphere below the main inversion base, rather than solely near the surface.

4 Discussion

4.1 Surface radiative balance

4.1.1 Shortwave

613 The small $SW_\downarrow$ biases exhibited by the standard UM configurations concurrent with a more significant SW_{net} bias indicate that
614 the modelled surface albedo is likely too low. While the observed albedo may be biased high due to its calculation from a
615 spatially small sample of sea ice (directly surrounding the ship), the UM surface albedo parameterisation has previously been
616 shown to be too low in the high Arctic (**Birch et al., 2009; 2012**). The temperature and albedo limits used in the standard
617 Regional Atmosphere parameterisation have been increased since **Birch et al. (2009, 2012)**; however, it is clear from **Fig. 2**
618 that the snow-on-sea-ice parameterisation limits tested here with ECMWF_IFS and UM_CASIM-100 (currently used in the
619 Global Atmosphere model configuration) produce a better comparison with our high Arctic measurements.

4.1.2 Longwave

621 The root of the LW error in each simulation is likely the >90% liquid-dominated low cloud occurrence which is not
622 representative of the observations (**Fig. 4b**). This problem has been previously identified in the high Arctic with both models
623 used in this study (**Birch et al., 2012**; **Sotiropoulou et al., 2016**) and recent model improvements/microphysical changes have
624 not sufficiently improved model performance in this regard. The positive LW biases are consistent with the too-warm surface
625 T biases in all simulations (**Figs. 9, 10**), which is also consistent with previous findings with both models (**Birch et al., 2009**;
626 **Sotiropoulou et al., 2016**) and with the ERA-Interim reanalysis product (**Jakobson et al., 2012**; **Wesslén et al., 2014**). **Figures**
627 **9** and **10** suggest that the UM simulations are perhaps better at capturing the near-surface T over the freeze, while ECMWF_IFS
628 consistently has a warm surface bias regardless of season. **Tjernström et al. (2021)** suggest that surface is actually warmed by
629 the atmosphere in the IFS, not the opposite, as indicated by the enhanced downward sensible heat flux, in combination with
630 diminished $SW_\downarrow$ with comparison to observations.

631 Given these results, we suggest that excessive cloudiness is likely a contributing factor to the warm surface bias in all
632 simulations. In particular, it is noteworthy that UM_CASIM-100 performs most poorly of the UM simulations. This result is
633 disappointing given the improvement of UM_CASIM-100 over the standard Regional Atmosphere configurations in both $SW_\downarrow$
634 and SW_{net}. Including CASIM aerosol processing through wet scavenging – thus enabling cloud dissipation (e.g., **Stevens et**
635 **al., 2017**) – may rectify this issue, or the representation of prognostic ice nucleating particles in place of a simple diagnostic
636 relationship between temperature and cloud ice number concentrations (e.g., **Varma et al., 2021**). These pathways will be
637 explored in future work; however, it is highly likely that other meteorological factors and incorrect model processes are
638 contributing to this warm surface bias across all of our simulations, in addition to cloudiness.

Atmospheric
Chemistry
and Physics
Discussions

4.2   Lower troposphere

4.2.1   Temperature

Temperature biases are strongest within the lowest 3 km of our model domains (**Figs. 9, 13**); this is also the altitude range over which the models overestimate cloud occurrence. With too much cloud, cloud top radiative cooling likely lowers the temperature too efficiently; this, coupled with incorrect cloud positioning (e.g., period 3; **Fig 13**), gives a cold bias with respect to our observations, above the observed main capping inversion. Where the liquid (and ice) phase is modelled more effectively – e.g., ECMWF_IFS during period 6 (**Fig. 13**) – the associated median biases are smaller ($< \pm 1$ K), supporting this conclusion.

The dipole in T errors shown in **Figs. 8** and **9**, with a positive bias towards the surface below a negative bias between 0.5 and 3 km, suggests that heat and moisture are not being sufficiently transported upwards from the surface or downwards from cooling at cloud top. This T bias is present in all simulations during both the melt and freeze periods and could result from the models failing to reproduce the structure of more than one strong observed inversion, instead exhibiting comparatively smooth T profiles. As shown by **Fig. 14**, low-altitude T inversions are often overestimated by the models, particularly during the melt period (not shown): this too-deep surface mixed layer likely results in incorrect cloud placement, leading to thermodynamic model-observation biases on a 1-to-1 comparison. Our scaled thermodynamic analysis (**Figs. 14, 15**) indicates that, while the models are often incorrectly placing the temperature inversion (consistent with previous findings; **Birch et al., 2012**), the relationship between too much cloud and strong negative T biases suggested by **Fig. 13** appears to be robust under these scaled height adjustments by the inversion base, and that the simulations are still largely too warm below the inversion.

Both **Sotiropoulou et al. (2016)** and **Tjernström et al. (2021)** found a similar vertical structure of the temperature biases with the IFS model, with positive biases within the lower 0.5 km of the atmosphere and a consistent cold bias present around 1 km. **Tjernström et al. (2021)** found that this cold bias intensifies with time during 3-day forecasts, indicating that it is made worse by processes within the model. They hypothesised the mid-level convection parameterisation triggering too-efficiently within the IFS could be transporting water vapour out of the BL, resulting in too much condensation to form cloud. While our UM LAM simulations do not employ such a convection scheme, the global driving model does: this parameterisation acts in addition to shallow convection, e.g., representing convection in mid-latitude storms, or above the BL where the surface layer is stable. Given the apparent close relationship between the biases exhibited by the LAM and global model **(Fig. 9, 11),** we conducted a short 6-day test with the global mid-level convection scheme switched off; this test caused over an order of magnitude increase in the UM_GLM cold biases shown in **Fig. 11** (not shown). Given the short duration and extreme "on-off" nature of this test, these results do not provide conclusive evidence that the mid-level convection scheme is not contributing to the thermodynamic biases shown here. Further investigation into vertical transport and mixing of scalars (temperature, moisture and clouds) is needed to confirm the origin of these thermodynamic biases. Specifically, more investigation into vertical transport and mixing of scalars (temperature, moisture and clouds) is needed; however, this investigation is beyond the scope of this paper.

4.2.2   Moisture

During the melt period, our results indicate that the UM is particularly moist throughout much of the troposphere (**Fig. 9b**), suggesting that the melting ice is enabling a too-great moisture source from the surface to the atmosphere. However, this

Atmospheric
Chemistry
and Physics
Discussions

Open Access

EGU

673   tropospheric bias appears to be rectified during the freeze, while the surface bias remains (**Fig. 9d**); therefore, the hypothesised

674   melting ice source is likely not the only contributor of this moisture bias. Latent heat fluxes measured during the expedition

675   indicate no significant change between the melt and freeze periods (not shown); therefore, the hypothesised increased moisture

676   flux during the melt is unlikely.

677   Given the close relationship between our UM LAM and global model biases (**Figs. 9, 11**), increased poleward moisture

678   transport introduced at the lateral boundary conditions from the mid-latitudes could partly explain these biases. This

679   phenomenon has been previously identified to be a consequence of climate change and may promote increased cloudiness in

680   the polar regions (e.g., **Held and Soden, 2006**; **Vavrus et al., 2009; Allen et al., 2012; Bender et al., 2012**). The moist surface

681   bias is also present over both the melt and freeze in ECMWF_IFS; however, ECMWF_IFS is routinely too dry between 0.5

682   and 4 km, in contrast to the UM. Instead, the IFS traps too much moisture in the lowest 0.5 km, suggesting that the upward

683   transport of moisture may be insufficient, the cloud sink above 0.5 km is too great, or there are consistent biases introduced via

684   assimilation of data other than the radiosonde data (e.g., satellite).

685   The moist bias exhibited by ECMWF_IFS towards the surface has previously been highlighted by **Sotiropoulou et al. (2016)**,

686   who suggested that this problem may explain why this model struggles to reproduce humidity inversions above the BL. There

687   are instances where negative T biases coincide with negative q biases at altitudes just above the main temperature inversion

688   (for example, at 27 Aug; **Fig. 9**). Moisture inversions have often been observed during the Arctic summertime (**Sedlar et al.,**

689   **2012; Nygård et al., 2014**); ECMWF_IFS fails to reproduce such inversions observed during AO2018. This dry bias above

690   the observed capping inversion around 27 Aug is not as strong in the UM simulations, but the UM does successfully reproduce

691   a small humidity inversion.

692   **4.3   Cloud macro- and microphysics**

693   The UM simulations have >98% cloud occurrence around 0.2 km over all four periods. Reduced $SW_\downarrow$ biases with respect to the

694   standard Regional Atmosphere configurations indicate that UM_CASIM-100 does improve agreement with our high Arctic

695   observations (**Table 4**). In particular, the comparison between UM_CASIM-100 and UM_RA2M (which use the same large-

696   scale cloud scheme and differ only in their representation of resolved cloud physics) shows that the new CASIM scheme

697   reproduces the observed Arctic clouds better on a microphysical level.

698   The ice phase differs more between the models than the liquid phase, likely due to its strong relationship with temperature:

699   UM_RA2M and UM_RA2T use the **Fletcher (1962)** parameterisation for primary ice formation, while ECMWF_IFS uses

700   **Meyers et al. (1992)** and UM_CASIM-100 uses **Cooper (1986)**. Each of these parameterisations is inherently temperature-

701   dependent, with **Meyers et al., (1992)** producing the largest ice number concentration, and **Fletcher (1962)** producing the

702   smallest. Given that each simulation does not reproduce the observed temperature profile well below 3 km, the onset of ice

703   nucleation (occurring below a threshold of –10 °C in the UM, for example) will be affected. If ice production is triggered

704   prematurely, cloud liquid properties should be dampened via the Wegener-Bergeron-Findeisen mechanism; evidence of this

705   can be seen in UM_RA2T during period 3, where an overestimation of ice below 2 km corresponds with a smaller mean LWC

706   than the other simulations (**Fig. 13b—c**).

707 When considering the drift as a whole, IWC is overestimated by all simulations (except UM_RA2M) below 1.5 km, where our

708 T and q biases are at their greatest. To test whether the method of parameterising primary ice itself has any effect on these

709 biases, we used the **Fletcher (1962)**, **Cooper (1986)**, and **Meyers et al. (1992)** parameterisations over a short test period within

710 the CASIM framework; however, we found little difference in the tropospheric ice with the different parameterisation methods

711 (**Fig. S8**). Given the spread in IWC results shown here, further investigation into the best methods to represent primary ice

712 production in such global and NWP models should be considered in future, with specific focus on employing prognostic ice

713 nucleating particles (similar to the CCN used here in UM_CASIM-100) to facilitate ice formation rather than simply using a

714 temperature-dependent function (e.g., **Varma et al., 2021**). The primary ice parameterisations used here do affect the clouds

715 modelled – for example, with a large spread in modelled IWC aloft during period 3 (**Fig. 13c**) – and a more realistic

716 representation of the ice phase would likely contribute to improved cloud liquid properties. In particular, it is likely that ice

717 formation at warm supercooled temperatures (> –10 °C) will be of importance given the overestimated dominance of cloud

718 liquid at low altitude in our simulations.

719 Below 3 km, the mean modelled LWC often exceeds the observed value, with better agreement between 0.15 and 1 km than

720 between 1 and 3 km (**Fig. 4, 13**). This overestimation of cloud liquid is also evident from the LWP data, with each simulation

721 exhibiting a greater LWP than was measured (**Fig. 5**) when averaged over our meteorological periods (not shown). At first

722 glance, the simulations agree reasonably well with measurements (**Fig. 5a**), but this subtle overestimation is clear from the

723 PDFs (**Fig. 5b**). The exception to this is UM_RA2T; this is the only simulation which often underestimates LWP, due to its

724 increased cloud ice mass within the lower troposphere in comparison to the other simulations (**Fig. 4d**). For example, the mean

725 measured LWP during period 3 is 122.8 g m^{-2}, yet UM_RA2T only produces 70.4 g m^{-2}. In contrast, UM_RA2T reproduces

726 the mean measured LWP well during period 6 (48.5 g m^{-2} measured versus 43.2 g m^{-2} modelled), with agreement improving

727 with time throughout the drift. This efficient ice-producing simulation suggests that the ice phase influences cloud properties

728 as time progresses more so in reality, while the other UM cases, with less dominant ice, retain too much liquid in comparison

729 to the measurements. To an extent, ECMWF_IFS also behaves in this way, retaining too much cloud liquid; however, it

730 performs much better than UM_RA2M and UM_CASIM-100 in reproducing the mean LWC and IWC during period 6 (**Fig.**

731 **13g, h**).

732 These simulations suggest that the model development community has effectively gone too far with the reduction of the ice

733 phase in central Arctic mixed-phase clouds. The surface LW balance is positively biased, and these excessive low-level clouds

734 are a contributing factor: by enabling too much liquid to form, and restricting the ice too efficiently, these clouds efficiently

735 absorb and re-emit upwelling LW radiation back towards the surface. Our results show that we have made great improvements

736 in the SW, driven by the improvements we have made to our cloud physics representation in these models (in addition to a

737 better estimation of the surface albedo). However, the too-consistent cloud cover coupled with too much cloud liquid is

738 hampering our model capability, and further developments (such as the inclusion of representative CCN and INP inputs to

739 double-moment cloud schemes to facilitate cloud dissipation) will likely go some way to tackle this issue.

Atmospheric
Chemistry
and Physics
Discussions

Open Access

EGU

740 **5** **Conclusions**

741 Model simulations with the Met Office UM and the ECMWF IFS were compared with observations made during the Arctic

742 Ocean 2018 expedition to evaluate model performance in the high Arctic, with particular focus on modelled clouds and the

743 surface radiative balance. Four key simulations were considered: a global configuration with the IFS and three nested

744 configurations with the UM (each using different combinations of large-scale cloud and microphysics schemes yet driven by

745 the same global model setup). These four simulations were compared with observations by using Cloudnet to build model-

746 comparable cloud fractions and water contents and thus identify consistent process weaknesses between the model

747 configurations chosen.

748 We found that key issues identified in previous studies, such as positive near-surface temperature biases (**Sotiropoulou et al.,**

749 **2016**), remain problems in recent releases of both the UM and IFS. Modelled BLs are often too deep (**Fig. 9, 14**), particularly

750 during the melt period, and thermodynamic biases, cloud occurrence, and cloud microphysics are consequently in poor

751 agreement with observations below 3 km. Excessive low-cloud occurrence is prevalent in both models (**Fig. 3**) and no

752 simulation adequately reproduces cloud-free periods and associated increases in longwave net emission (**Fig. 2**), consistent

753 with previous UM and IFS evaluations in the Arctic (**Birch et al., 2012**; **Sotiropoulou et al., 2016**). Strong negative

754 temperature biases (**Figs. 8, 9, 13**) coincide with too-frequent liquid-dominated cloud layers (**Fig. 13a, b, f, g**), likely associated

755 with over productive cloud-top radiative cooling in the models. Cloud liquid and ice water contents, especially below 1 km,

756 were within an order of magnitude of our observations (**Fig. 4**), but clouds occurred too frequently, contained too much liquid

757 between 1 and 3 km, and were often at too-high an altitude (**Fig. 13a—c**).

758 Radiative interactions are in better agreement with observations and all models capture the observed distribution of SW_{net} and

759 LW_{net} better during the sea ice freeze period in comparison to the melt period (**Fig. 9**). Improved radiative interactions and

760 thermodynamic biases during the freeze can be linked with improved agreement of cloud occurrence and microphysics **(Fig.**

761 **13, S7).** We found that the surface albedo in each model configuration is underestimated with respect to observational estimates

762 (see **Supporting Information**), but this is unsurprising given the models are representing an average albedo over a 1.5 / 9 km

763 grid box while our observed estimates are from the area immediately surrounding the ship. Updating the surface albedo

764 parameterisation limits used within the UM Regional Atmosphere configurations (UM_RA2M/UM_RA2T) to those used in

765 the Global Atmosphere GA6.0/6.1 configuration (UM_CASIM-100) greatly improves our surface albedo comparison with

766 observational estimates (see **Supporting Information**) and thus contributes to the good comparison of UM_CASIM-100 with

767 measured shortwave radiation data.

768 We propose that four factors are important to failings in our model simulations:

769   1. The choice and use of large-scale cloud schemes at high resolution:

770   o   Both the UM and IFS poorly capture Cloudnet-calculated cloud fractions from observations over Aug—Sep

771   2018 in the central Arctic, particularly at altitudes between 2 km and 8 km (**Fig. 4a**). Building a comparable

772   mask based on TWC shows that the cloud modelled aloft is actually in good agreement with observations

773   (**Fig. 4b**), while highlighting that the over prediction of cloud occurrence below 3 km is in fact much worse

774   than suggested by the C_V comparison. As such, we suggest that cloud fractions should not be used in isolation

775   as a model comparison metric over the Arctic as models represent this parameter differently at the present

776   time (as detailed in the **Supporting Information**).

777   2.   The cloud microphysics scheme chosen to represent resolved clouds:

778   ○   UM_CASIM-100 performs best in terms of SW_{net} (**Fig. 2**, **Table 4**), but it struggles to capture cloud-free

779   episodes, thus producing a LW_{net} PDF which is too narrow in comparison to our measurements.

780   ○   ECMWF_IFS shares the too-narrow LW_{net} PDF of UM_CASIM-100; however, it often produces a IWC in

781   reasonable agreement with observations, and its mean LWC profile does agree particularly well with the

782   observations at times (e.g., period 6; **Fig. 13g**).

783   ○   Of the UM simulations considered, UM_CASIM-100 is in best agreement with both ECMWF_IFS and

784   observations in terms of net radiation, SW_{net} and $SW_\downarrow$. This improved radiative agreement can be linked to

785   its better cloud microphysical agreement with our Cloudnet-derived cloud liquid water content over the

786   standard Regional Atmosphere configurations (**Figs. 4, 5, 13**); however, UM_CASIM-100 produces even

787   poorer cloud fractions aloft than either UM_RA2M or UM_RA2T.

788   3.   Representative CCN concentrations, and thus droplet number concentrations, as a function of altitude:

789   ○   Representative CCN concentrations in UM_CASIM-AeroProf somewhat improves the overestimation of q_{liq}

790   within low level clouds in UM_CASIM-100. However, the q_{liq} decrease is not sufficient to trigger an

791   increase in liquid precipitation, which would thus decrease cloud lifetime, so the modelled C_V is essentially

792   unchanged (**Fig. 7**). Crucially, thermodynamic biases with respect to observations are not improved through

793   this enhanced complexity (**Fig. 10**), highlighting that these biases may not be fixed by a more comprehensive

794   representation of cloud physics. Further work is required, with the inclusion of wet scavenging of aerosols

795   and prognostic INP, to rule out whether such processes could improve the model biases over and above the

796   inclusion of representative aerosol concentrations alone.

797   4.   The global model analyses used to produce boundary conditions for high resolution nests:

798   ○   The thermodynamic biases identified in our models differ only a little between the UM simulations despite

799   differences in their cloud configurations. Comparisons with the global model show that the biases within the

800   LAM are largely inherited from the global model and associated DA system (**Fig. 9, 11**); therefore, for LAM

801   configurations such as that tested here, we will not obtain the true benefit of more sophisticated cloud

802   microphysics schemes in NWP simulations until we address the large-scale biases in their driving models/DA

803   system.

804   While representative CCN concentrations are indeed important for properly reproducing Arctic cloud structure and its

805   consequential impact on the net surface radiation, our findings indicate that such representative cloud nuclei inputs still have

806   little impact on thermodynamic biases in the lower troposphere. For our given LAM configuration, we speculate that these

807   biases will always be inherited from the driving model/DA and will continually bias cloud formation processes and BL depth;

808   however, using an increased domain size, with the science area of interest as far from the lateral boundaries as possible, may

809   help to reduce the influence of the driving model/DA. The issue of inherited thermodynamic biases is concerning as both the

810   UM global model and IFS are both used within the community to drive NWP configurations of the same model (UM) or others

811 (IFS). For example, the IFS configuration tested here is similar to that used by ERA5; therefore, these biases could influence

812 future high Arctic NWP simulations if these reanalyses are used for initialisation.

813 Our recommendations are thus twofold. To improve our Arctic cloud modelling capability, we must continue to improve the

814 cloud physics description striving for an optimum complexity, such as the introduction of representative CCN concentrations

815 and double-moment cloud liquid illustrated here, in addition to the inclusion of prognostic INP and associated aerosol

816 processing mechanisms. However, we must concurrently address the overabundant occurrence of a too-well-mixed and too-

817 cloudy lower troposphere, and tackle the resultant thermodynamic biases, in our global driving models and their respective DA

818 systems.

819

820   **Appendix A   Cloudnet mishandling of fog data**

821   LWP measurements from the HATPRO microwave radiometer were used in this study; this instrument provides measurements

822   of microwave brightness temperatures, from which LWP is derived for the full atmospheric column above the instrument (here

823   located approximately 13 m above the surface). This includes measurement of liquid clouds at altitudes below the radar's first

824   range gate at 156 m. Fog periods occurred frequently throughout the expedition (**Vüllers et al., 2021**); therefore, we had several

825   instances where liquid fog was measured with the HATPRO with small quantities of liquid, or none, detected in the clouds

826   above (from lidar/radar).

827   Cloudnet calculates an offset to be deducted from the LWP time series dependent upon its categorisations of cloud to ensure

828   that liquid is partitioned throughout the cloud column only if liquid clouds were present. This offset is non-uniform, calculated

829   as a given fraction of the LWP signal on a daily basis, and is deducted from the LWP data to ensure liquid partitioning is

830   conducted correctly within the Cloudnet algorithm. Given the frequency of fog occurrence, this offset was often overestimated

831   and too much liquid data were removed, thus negatively impacting the LWP and LWC comparisons with our model simulations.

832   To rectify this problem, we removed the LWP offset calculation from the Cloudnet procedure, enabling all ingested data to be

833   used by Cloudnet. We then compared these adapted Cloudnet LWC data to a LWC calculated under an adiabatic assumption to

834   test whether the latter could be used as an approximation of the true LWC if there was not as much fog present during the

835   expedition. **Figure A1** shows this comparison using all data from the drift period and indicates that, by keeping all fog liquid

836   data in the time series, Cloudnet artificially partitions these data to liquid cloud layers identified by the lidar, leading to too

837   much liquid in clouds within the lowest 1 km of the atmosphere (with comparison to the adiabatic profile). These data also

838   indicate that we can safely use the adiabatic LWC as this artificial liquid enhancement is confined to the lowest 1 km and does

839   not significantly affect the comparison for higher altitudes. Following these comparisons, we chose to include the adiabatic

840   LWC in our comparisons with model simulations to exclude the artificial enhancement of cloud liquid at low altitudes in our

841   measurement data.

842

843   **Tables**

Table 1: List of key abbreviations, acronyms, and parameters referenced in this study.	
Label used	**Description**
β	Lidar backscatter
BL	Boundary layer
C_V	Cloud fraction by volume, determined using Cloudnet
CASIM	Cloud-AeroSol Interacting Microphysics
CCN	Cloud condensation nuclei
ECMWF	European Centre for Medium-Range Weather Forecasting
GCM	General Circulation Model
IFS	Integrated Forecasting System
IWC	Cloud ice water content
LAM	Limited Area Model
LW_{net}	Net longwave radiation at surface
$LW_\downarrow$	Downwelling longwave radiation at surface
LWC	Cloud liquid water content
LWP	Liquid water path
N_d	Cloud droplet number concentration
q	Specific humidity
q_{ice}	Cloud ice mixing ratio
q_{liq}	Cloud liquid mixing ratio
R_{net}	Net total radiation at surface
RH	Relative humidity
RH_{crit}	Critical relative humidity for condensation in models
SW_{net}	Net shortwave radiation at surface
$SW_\downarrow$	Downwelling shortwave radiation at surface
θ_e	Equivalent potential temperature
T	Temperature
TWC	Total cloud water content
UM	Unified Model
Z_e	Radar reflectivity

Atmospheric
Chemistry
and Physics
Discussions

844

Table 2: Summary of the four model configurations to simulate cloud and thermodynamic conditions observed over the full AO2018 drift period in this study. Three additional simulations included for further investigation of results are listed in shaded sections below.

Simulation	Details	References
ECMWF_IFS	Cy46r1; cloud and large-scale precipitation following update to Cy36r4. Snow included in all cloud fraction and cloud ice water content analyses.	**Forbes and Ahlgrimm (2014)**
UM_CASIM-100	UM with LAM using CASIM scheme operating with 100 cm^{-3} accumulation mode aerosol particles over the full model column and across the entire LAM. Droplet activation: **Abdul-Razzak and Ghan (2000)**; primary ice formation: **Cooper (1986)**. Diagnostic cloud fraction and condensate from large-scale cloud (**Smith, 1990**) scheme.	**Smith (1990)**; **Hill et al. (2015)**; **Grosvenor et al., (2017)**; **Kupiszewski et al. (2013)**
UM_RA2T	UM with LAM operating with the tropical regional atmosphere configuration (RA2T). Prognostic cloud and prognostic condensate (PC2) cloud scheme used with cloud microphysics based on **Wilson and Ballard (1999)**.	**Wilson and Ballard (1999)**; **Wilson et al. (2008)**; **Bush et al. (2020)**
UM_RA2M	UM with LAM operating with the mid-latitude regional atmosphere configuration (RA2M). **Wilson and Ballard (1999)** cloud microphysics scheme with diagnostic cloud fraction and condensate from large-scale cloud (**Smith, 1990**) scheme.	**Smith (1990)**; **Wilson and Ballard (1999)**; **Bush et al. (2020)**
UM_GLM	UM global model operating a N768 resolution (corresponding to approximately 17 km at the mid-latitudes) using the Global Atmosphere 6.1 configuration with a rotated pole. Uses 70 quadratically-spaced vertical levels up to 80 km with PC2 large-scale cloud and cloud microphysics based on **Wilson and Ballard (1999)**. Data over the full drift period are included to contextualise thermodynamic profiles extracted from the UM LAMs.	**Walters et al. (2017)**; **Wilson et al. (2008)**; **Wilson and Ballard (1999)**
UM_CASIM-AeroProf	As UM_CASIM-100, except day-averaged soluble coarse- and accumulation-mode concentrations from UKCA are input in place of the constant profile (see **Supporting Information** for details) to indicate role of realistic aerosol number concentrations.	**Morgenstern et al. (2009)**; **O'Connor et al. (2014)**; **Mann et al. (2010)**

| UM_RA2M-ERAI-GLM | As UM_RA2M LAM configuration, except using ERA-Interim data to initialise the UM global model instead of standard global start dumps. Data only included from a short subset of the drift period (31 Aug to 5 Sep) for further analysis of temperature and moisture profiles. | **Dee et al. (2011)** |

845

Table 3: Summary of cloud microphysical process representation in each simulation setup. Chosen processes are highlighted as key differences between the schemes. k = model level; Z = altitude.

	Model simulation			
Physical process	**ECMWF_IFS**	**UM_CASIM-100**	**UM_RA2T**	**UM_RA2M**
Prognostic cloud variables	Cloud fraction, vapour, cloud liquid, cloud ice, rain, and snow (single moment)	Vapour, cloud liquid, cloud ice, graupel, rain, and snow mixing ratios and number concentrations (double moment)	Liquid, ice, and total cloud fractions; vapour, cloud liquid, cloud ice (all ice, includes snow), and rain (single moment).	Vapour, cloud liquid, cloud ice (all ice, includes snow), and rain (single moment).
Large-scale cloud fraction (*described in*)	Prognostic (**Tiedtke, 1993**)	Diagnostic (**Smith, 1990**)	Prognostic (**Wilson et al., 2008**)	Diagnostic (**Smith, 1990**)
Droplet number concentration	Diagnostic. Wind-speed dependent function for radiation calculations (following **Martin et al., 1994**). For auto-conversion, diagnosed by land-surface mask (ocean surface, fixed); 50cm^{-3}.	Prognostic; **Abdul-Razzak and Ghan (2000)**, referencing an accumulation mode aerosol profile of 100cm^{-3} at all Z.	Diagnosed by land-surface mask (ocean surface, fixed); 100 cm^{-3}. Tapered to 50 cm^{-3} at $Z \leq 50$ m from 150 m.	Diagnosed by land-surface mask (ocean surface, fixed); 100 cm^{-3}. Tapered to 50 cm^{-3} at $Z \leq 50$ m from 150 m.
Critical grid-box mean RH for condensation	$RH_{crit} = 0.8$, increasing towards the BL as a function of height.	0.96 at the surface and decreases monotonically upwards to 0.80 at 0.85 km, above which it remains constant with altitude (k >= 15) (**Grosvenor et al., 2017**)	0.96 at the surface and decreases monotonically upwards to 0.80 at 0.85 km, above which it remains constant with altitude (k >= 15) (**Grosvenor et al., 2017**)	0.96 at the surface and decreases monotonically upwards to 0.80 at 0.85 km, above which it remains constant with altitude (k >= 15) (**Grosvenor et al., 2017**)

846

847

Table 4: Mean surface radiation biases (model-observations) over periods 3—6, with mean measured values for reference. Observations included are hourly-integrated values for consistency with the models. All values are in W m^{-2}. Smallest biases are highlighted in bold.

Component		Observations	ECMWF_IFS	UM_CASIM-100	UM_RA2T	UM_RA2M
SW_{net}	P3	24.35	16.41	**6.09**	39.69	38.02
	P4	19.45	4.23	**0.85**	24.74	20.77
	P5	9.87	9.75	**7.88**	19.81	18.09
	P6	7.37	6.75	**5.11**	16.67	15.3
$SW_{\downarrow}$	P3	117.39	-20.4	-10.93	7.44	**4.08**
	P4	72.86	-6.84	**-1.49**	14.68	6.99
	P5	55.55	**0.9**	9.7	16.43	12.66
	P6	41.66	**-0.3**	-2.21	11.86	7.48
LW_{net}	P3	-21.38	**9.56**	16.32	10.48	10.71
	P4	-9.48	3.16	4.85	**2.71**	3.25
	P5	-11.77	-2.96	**1.89**	-3.77	-3.65
	P6	-13.22	**-0.27**	5.46	-3.6	-0.29
$LW_{\downarrow}$	P3	285.82	19.0	22.41	**18.42**	18.66
	P4	303.12	6.0	6.5	**5.04**	5.54
	P5	291.53	**0.31**	2.6	-2.25	-2.15
	P6	286.88	5.02	9.6	**0.28**	5.13

848

Atmospheric
Chemistry
and Physics
Discussions

849 **Figures**
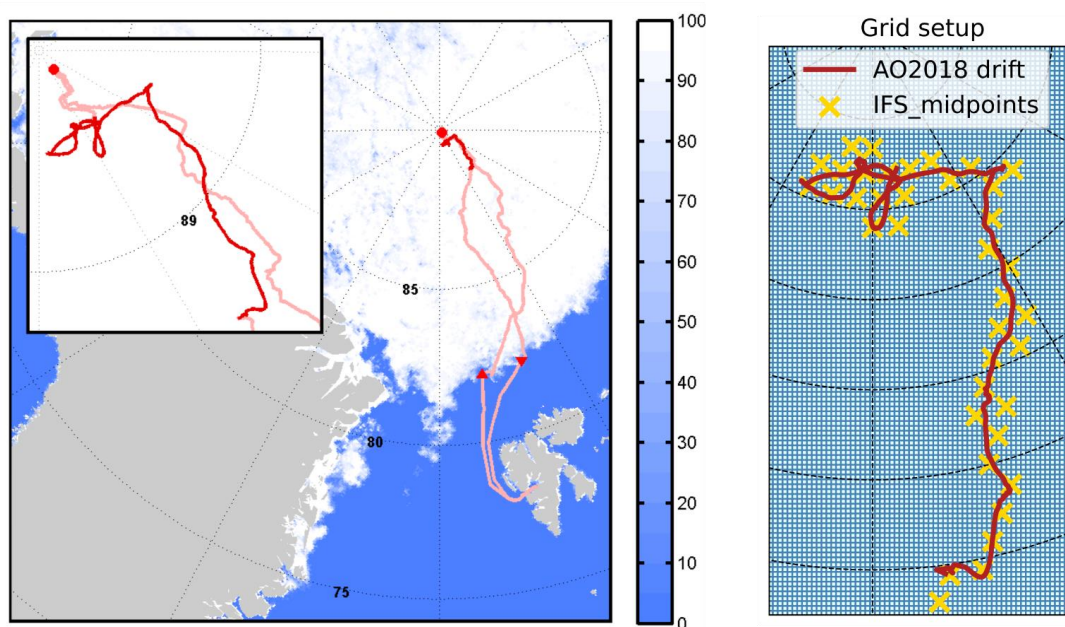
Figure 1: LHS: Map of cruise track and sea ice cover during AO2018 expedition from **Vüllers et al. (2021)**, with drift period (red) in inset. RHS: Ship position during the drift period (red), with grid outline for UM_CASIM-100, UM_RA2T, and UM_RA2M shown in blue and mid-points of ECMWF_IFS grid indicated by yellow crosses. Note grid size difference for illustrative purposes and not to scale: UM grid boxes are 1.5 × 1.5 km, IFS grid boxes are 9 × 9 km in size.

850
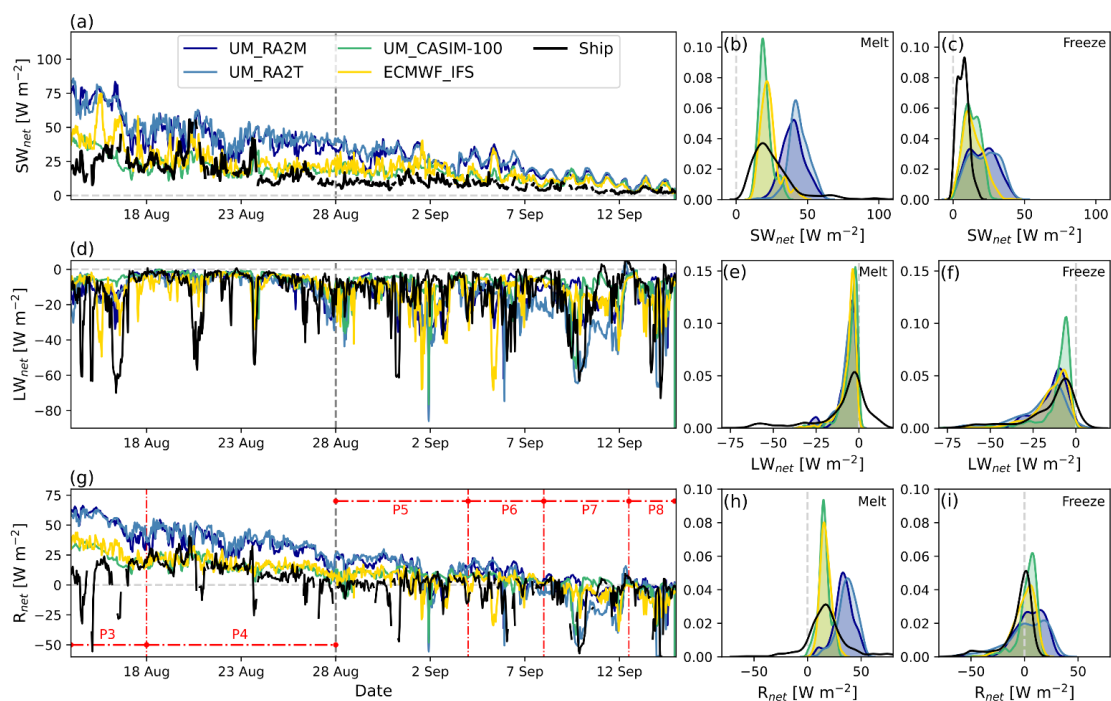
Figure 2: SW_{net}, LW_{net}, and R_{net} simulated by UM_RA2M (dark blue), UM_RA2T (light blue), UM_CASIM-100 (green), and ECMWF_IFS (yellow). Hourly-averaged measurements on board the ship (black) shown for comparison. LHS: timeseries; RHS: PDFs. PDFs are split between melting and freezing sea ice conditions using a threshold of 28 Aug as indicated by the grey vertical dashed line in panels (a), (d), and (g). Radiation terms are defined as positive downwards. Sub-periods used in subsequent sections are marked (red) in panel (g).

851
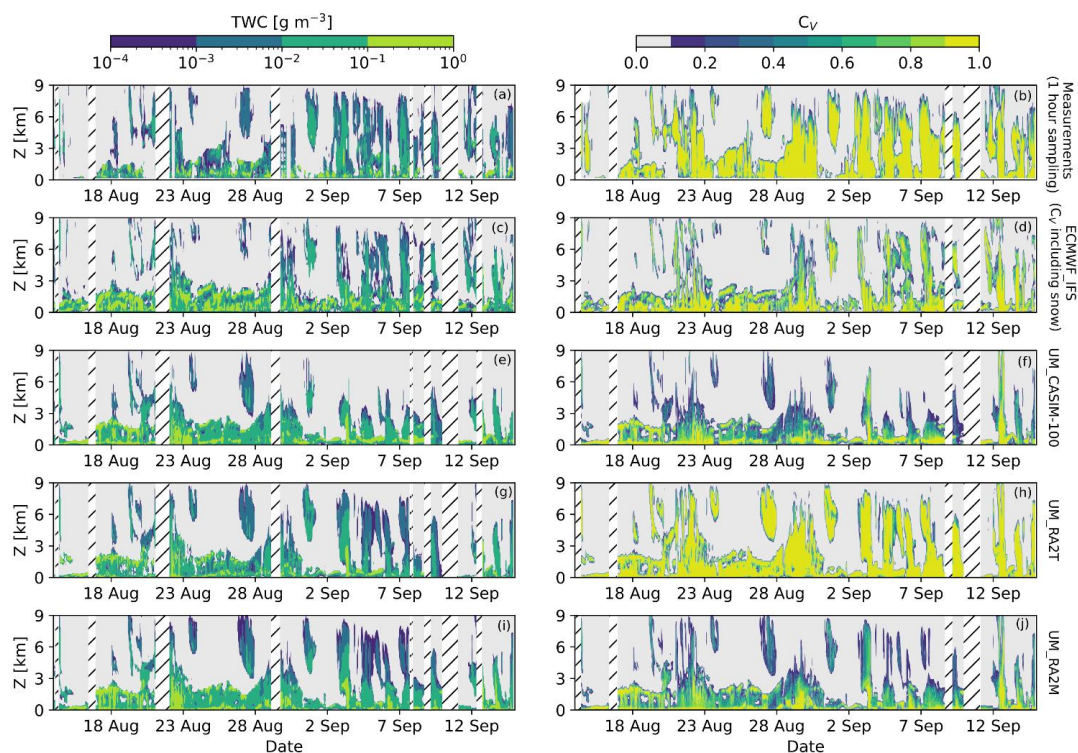
Figure 3: Total water content (*left, TWC*) and cloud fraction (*right, C_V*). (**a—b**) calculated from observations using Cloudnet and diagnosed by (**c—d**) ECMWF_IFS, (**e—f**) UM_CASIM-100, (**g—h**) UM_RA2T, and (**i—j**) UM_RA2M. Missing measurement data are indicated by hatched areas; times where data are missing from the observations are removed from the model data to provide a fair comparison. Missing data periods differ between the TWC and C_V products due to the different instrumentation requirements within Cloudnet for each.

852
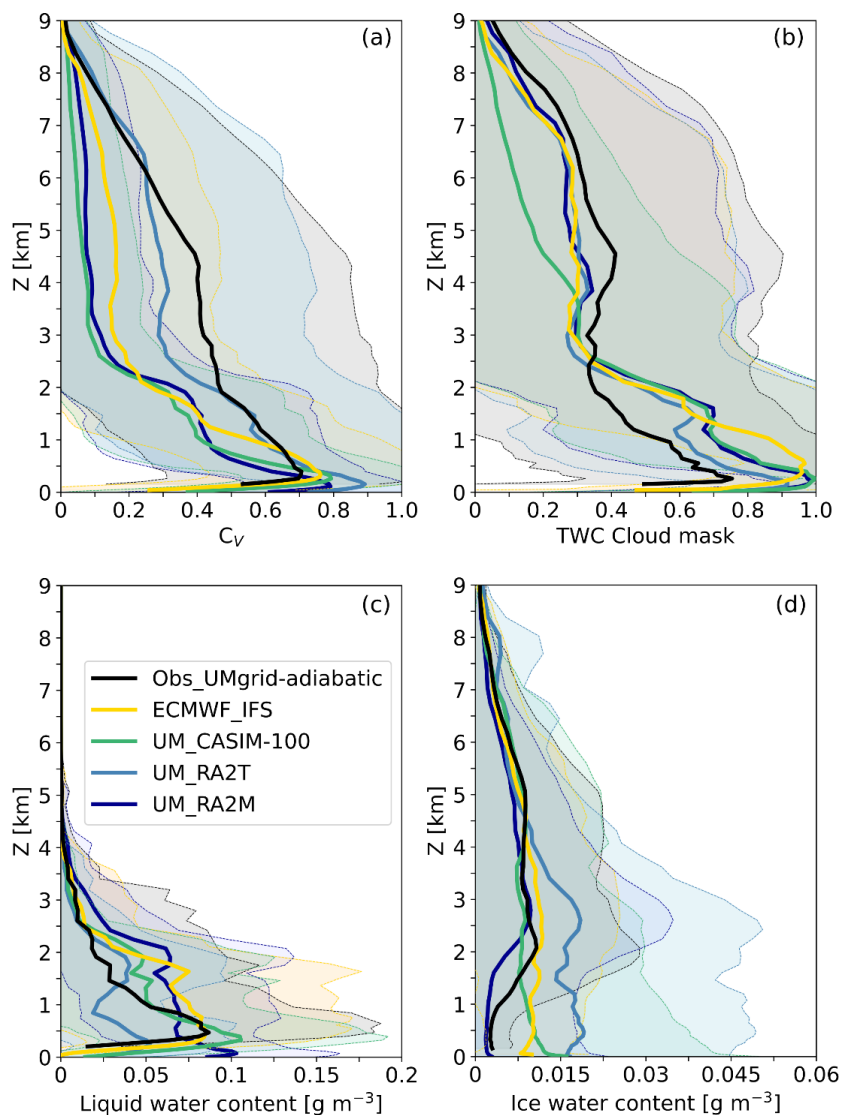
Figure 4: Comparison between (**a**) mean C_V observed (black, calculated using Cloudnet) and modelled (UM_RA2M = dark blue; ECMWF_IFS = yellow; UM_CASIM-100 = green; UM_RA2T = light blue) over the AO2018 drift period. (**b**) TWC cloud mask comparison, where masks are calculated using only in-cloud data as described in **Sect. 2.4**. (**c—d**) Same comparison for liquid and ice cloud water contents respectively, using in-cloud data only. LWC data from the observations are calculated using Cloudnet by assuming an adiabatic profile (see **Appendix A**). Lines indicate the mean profiles of each dataset, shaded areas depict ± one standard deviation from the mean. Uncertainties associated with the retrieval process are not shown.

853
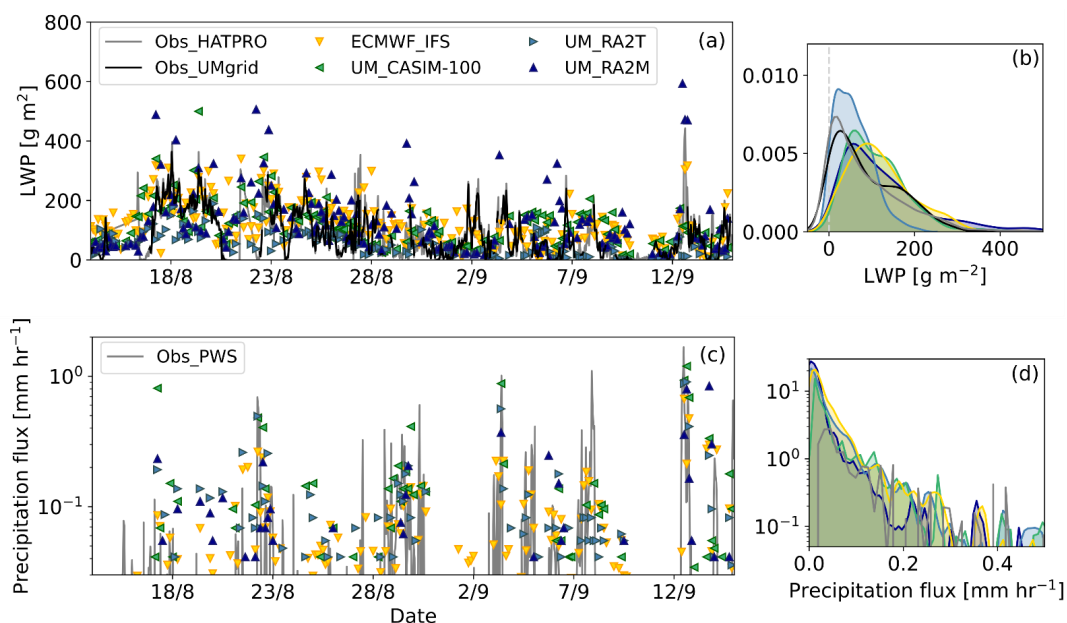
Figure 5: Timeseries of (**a—b**) liquid water path (*LWP*) and (**c—d**) total precipitation flux at the surface over the full drift period. (**a—b**) HATPRO measurements (grey) are included for comparison with the model data (coloured markers). *LWP* data averaged on to the UM grid by Cloudnet are shown in black (*Obs_UMgrid*). (**c—d**) Weather sensor (PWS) measurements of total precipitation from the 7th deck (grey) are included for comparison with model rain and snow fields. (**a, c**) model data shown every 3 hours for clarity; (**b, d**) all model data included for comparison.

854

855
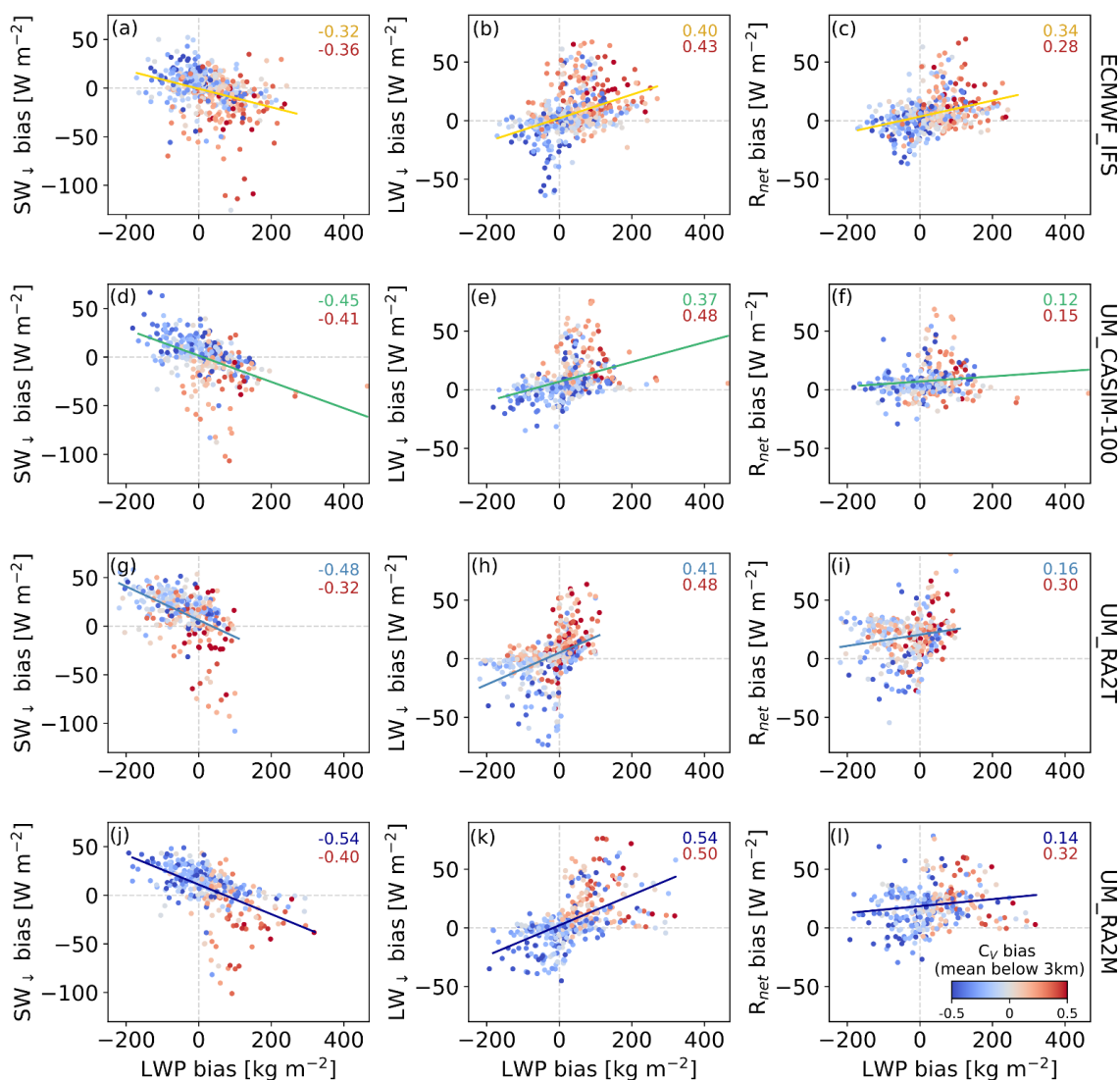
Figure 6: Model biases in radiation terms ($SW_\downarrow$ (*left*), $LW_\downarrow$ (*middle*), and R_{net} (*right*)), LWP, and C_V. Model-observation biases are calculated hourly for the radiation and LWP terms using measurements from the ship-based radiometers and HATPRO microwave radiometer, respectively. Shading: model-observation difference between mean C_V below 3 km, where model data below the height of the lowest radar range gate (156 m) is excluded from the comparison with observations. Correlation coefficients for the radiation-LWP (*top*) and radiation-C_V regressions (*bottom*) are noted in the top right of each panel.

856

857
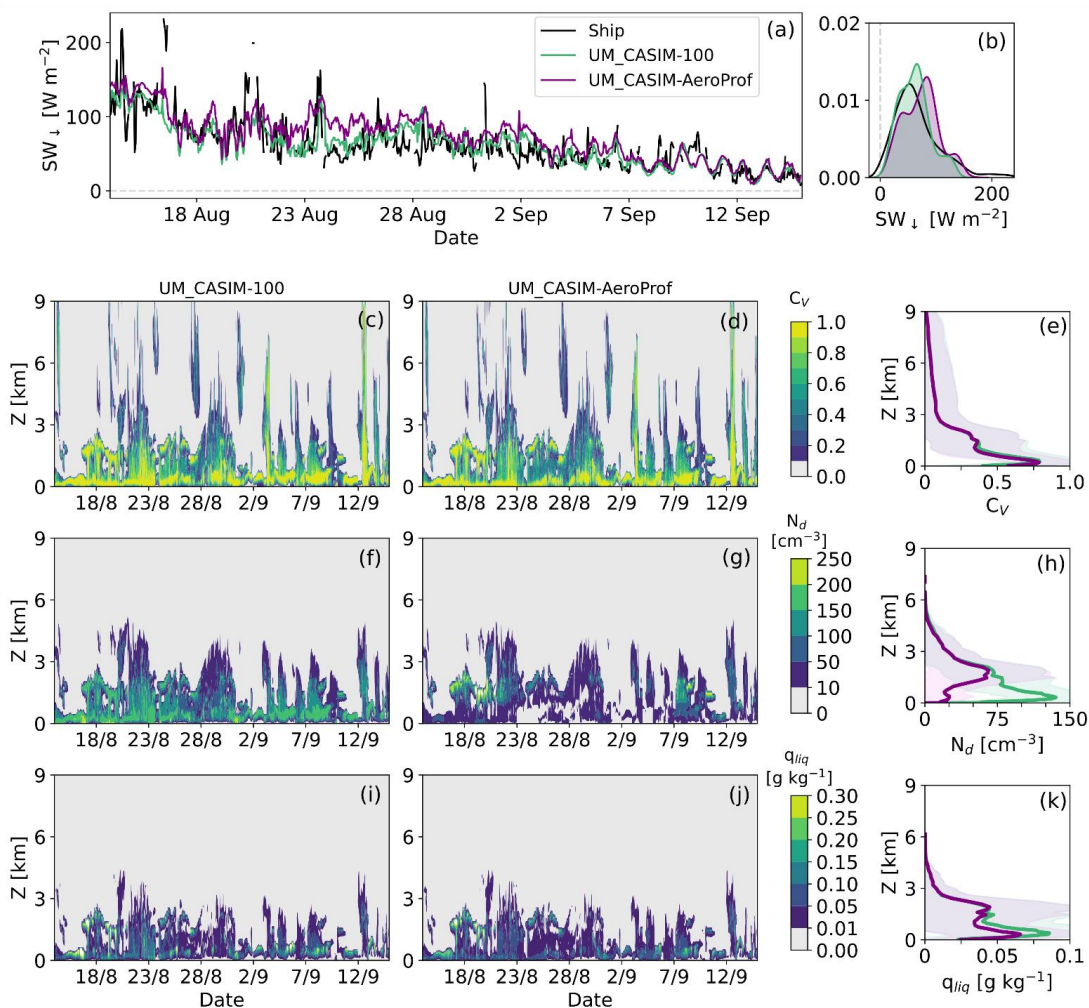
Figure 7: Comparison of UM_CASIM-100 and UM_CASIM-AeroProf, demonstrating the influence of representative aerosol concentrations on the modelled cloud structure. (**a—b**) downwelling shortwave radiation ($SW_\downarrow$) at the surface, with observations (black) shown for comparison; (**c—e**) C_V; (**f—h**) cloud droplet number concentration (N_d); and (**i—k**) liquid water mixing ratio (q_{liq}). (**c, f, i**): UM_CASIM-100; (**d, g, j**): UM_CASIM-AeroProf; (**e, h, k**): mean profiles with ± one standard deviation shown in shading. Radiative differences are only notable between 22 Aug and 27 Aug. Slight differences in q_{liq} and cloud fraction can also be identified during this period; for example, UM_CASIM-100 produces a larger cloud fraction below 2 km at 23 Aug

858

859
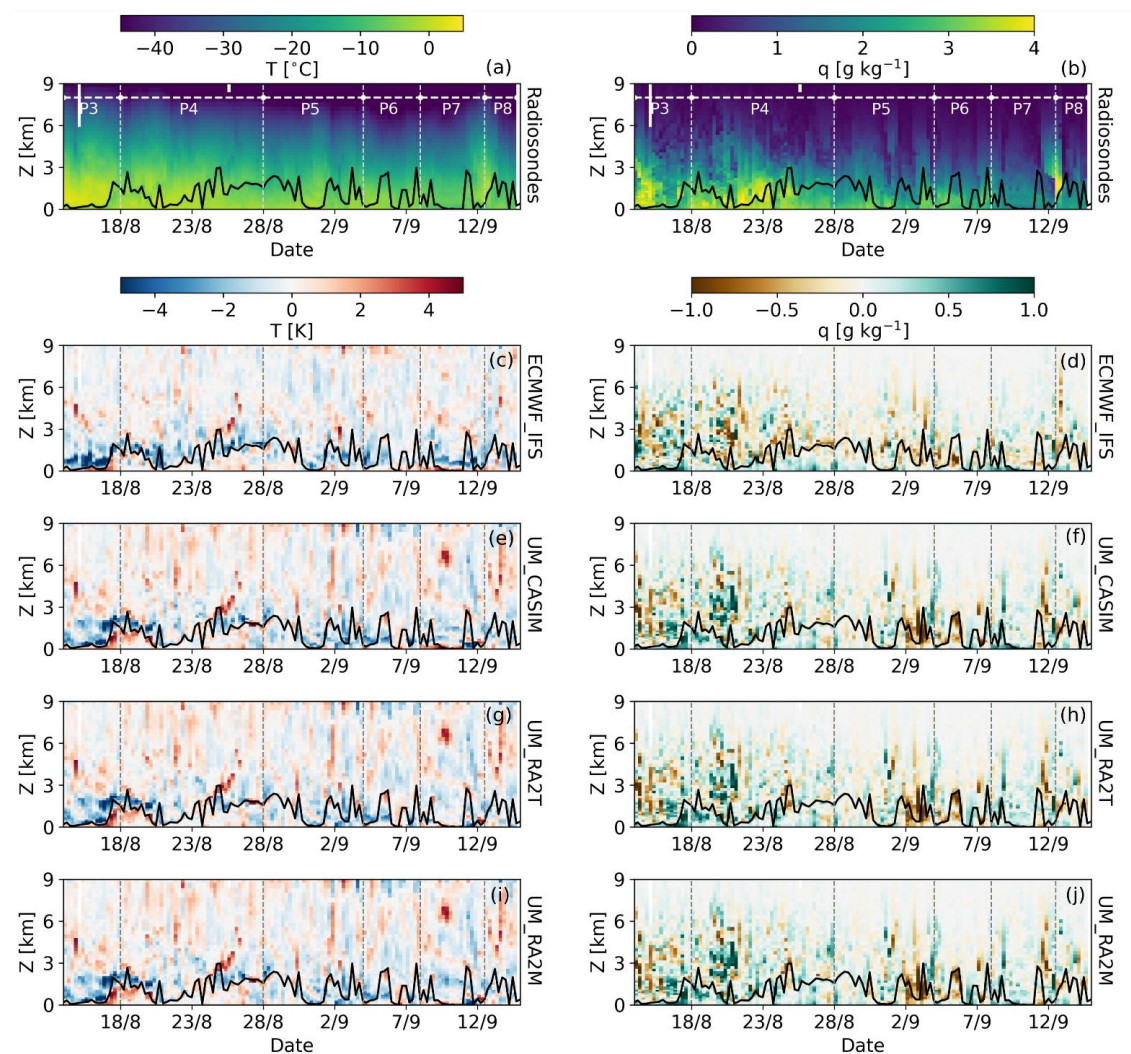
Figure 8: T (left) and q (right) measured by the radiosondes over the AO2018 drift period. (**a—b**): radiosonde data re-gridded to the UM vertical grid for model comparisons. (**c—d**): biases of IFS data, re-gridded to the UM vertical grid, with respect to observations. (**e—j**) UM_CASIM-100, UM_RA2T, and UM_RA2M biases, with no vertical re-gridding. The common vertical grid (from the UM) provides 50 vertical levels below 10 km, with 21 of these below 2 km. The black line in all panels depicts the altitude of the main inversion base as identified using the radiosonde measurements, and meteorological time periods with common characteristics are indicated in white (see **Vüllers et al., 2021** for details).

860
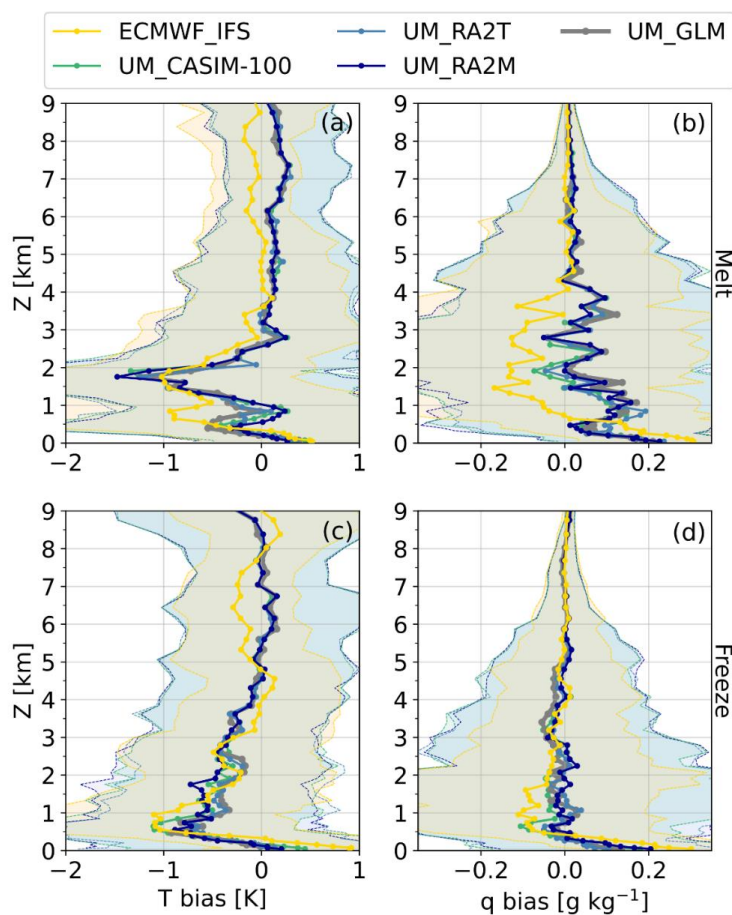
Figure 9: Median profiles of modelled (**a, c**) T and (**b, d**) q biases with respect to the radiosonde measurements over the sea ice melt (*top*) and freeze (*bottom*) periods (using 28 Aug as a threshold). Model data are coloured as previous (ECMWF_IFS: yellow; UM_CASIM-100: green; UM_RA2T: light blue; and UM_RA2M: dark blue) and ± one standard deviation shown to illustrate variability. Median anomalies from the UM global model (UM_GLM; grey) are also included for reference; variability is not shown for these data.

861

862
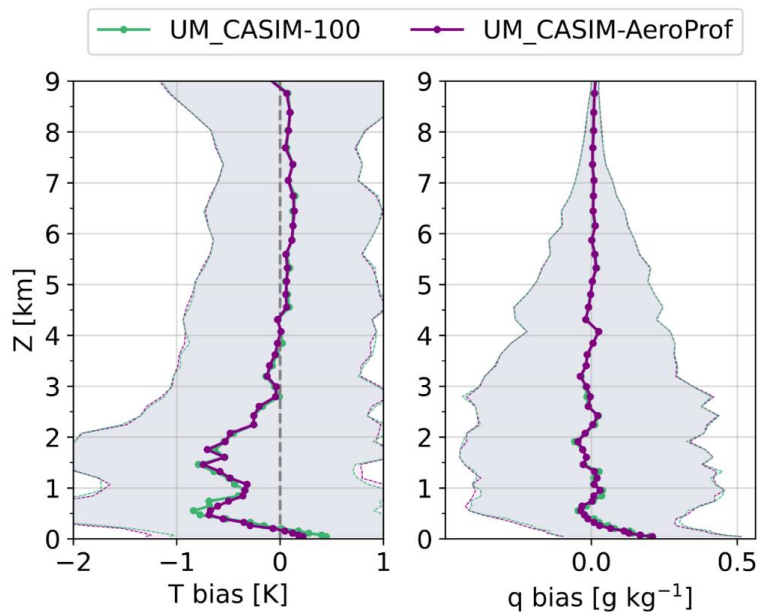
Figure 10: Temperature (*left*) and moisture (*right*) biases exhibited by the UM_CASIM-100 (green) and UM_CASIM-AeroProf (purple) simulations with respect to radiosonde measurements made over the entire drift period. ± one standard deviation shown in shading to illustrate variability.

863

864
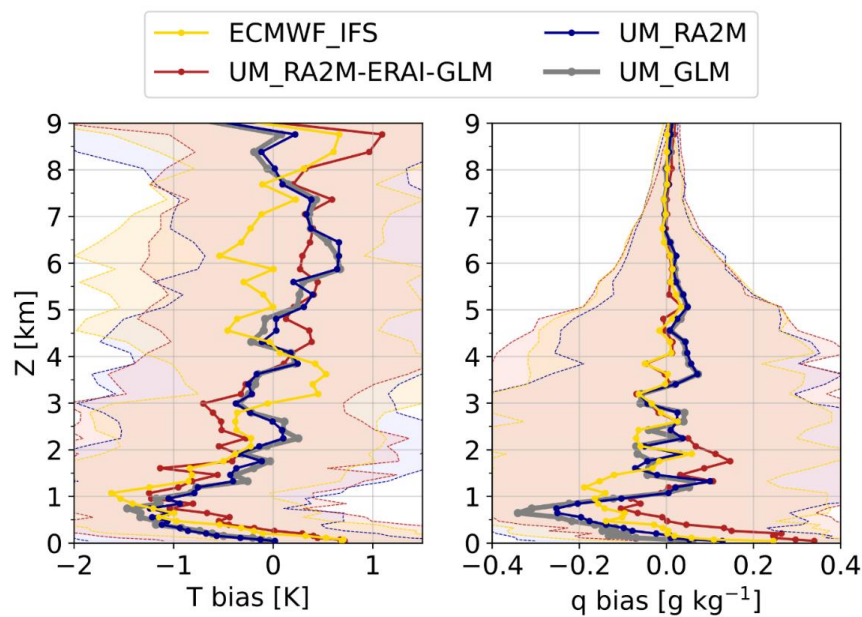
Figure 11: Median T and q biases from a subset of the drift (31 Aug to 5 Sep) for ECMWF_IFS (yellow), UM_RA2M-ERAI-GLM (red), UM_RA2M (dark blue), and UM_GLM (grey). UM_RA2M-ERAI-GLM biases follow ECMWF_IFS biases up to approximately 1 km, above which they largely behave more like the other UM cases. ± one standard deviation shown in shading to illustrate variability.

865

866

867
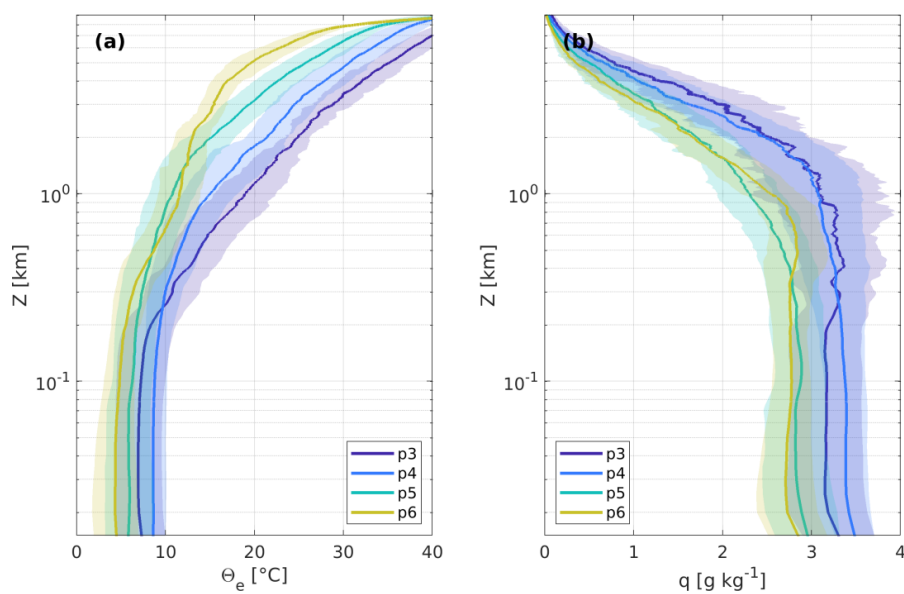
Figure 12: Mean profiles of (a) equivalent potential temperature (θ_e) and (b) q measured by radiosondes launched during periods 3—6 of the expedition, with ± one standard deviation shown in shading.

868
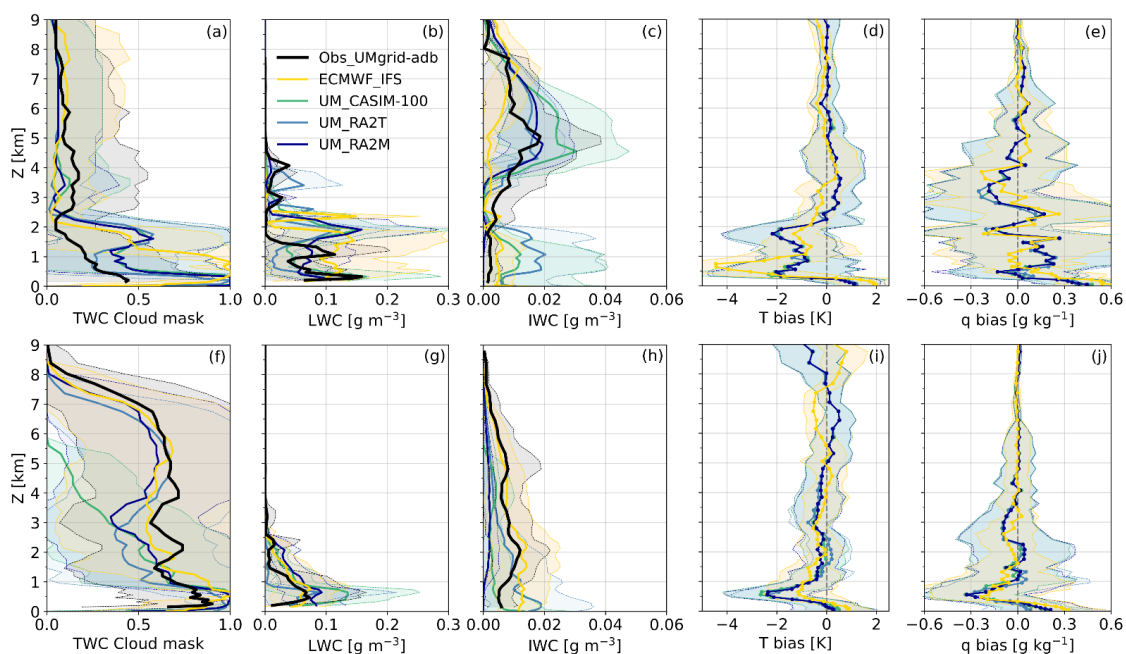
Figure 13: Comparison of mean cloud mask, LWC, and IWC profiles with median biases in T and q with respect to radiosondes for period 3 (**a—e**, *top row*) and period 6 (**f—j**, *bottom row*). Again, observed LWC calculated assuming adiabatic conditions using Cloudnet. ± one standard deviation shown in shading to illustrate variability.

869

870

871
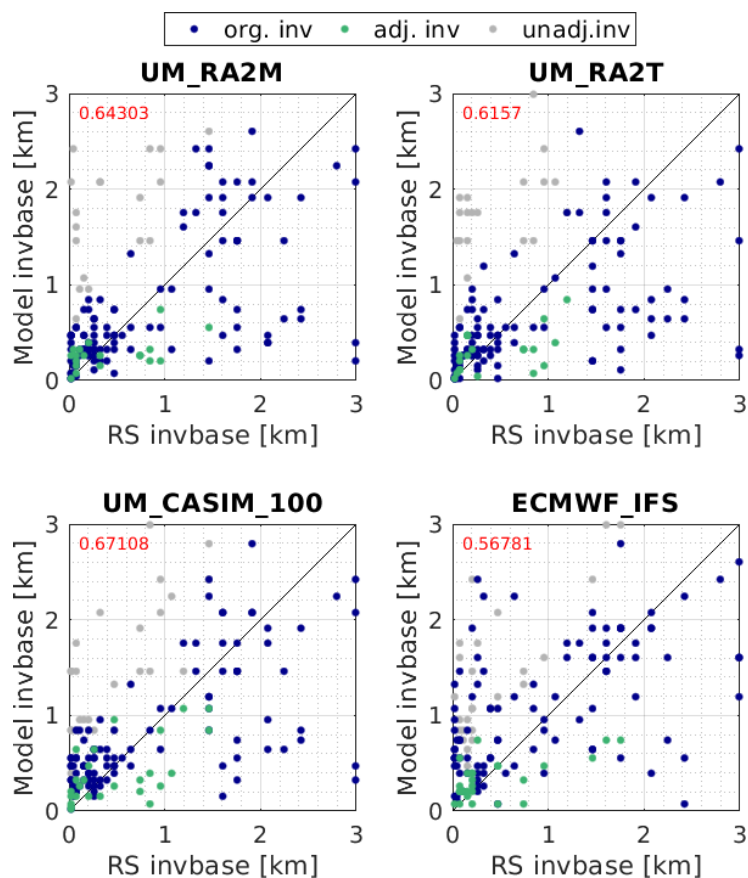
Figure 14: Model temperature inversion base as a function of the identified inversion base from radiosonde (RS) measurements. *org. inv*: strongest inversion below 3 km, identified following **Vüllers et al. (2021)**. *adj. inv:* where models exhibit a secondary weaker inversion at lower altitude in better agreement with identified radiosonde inversions, these identified inversions are adjusted accordingly. *unadj.inv*: unadjusted primary inversions, not used for further analysis and shown for reference only. Correlation coefficients are quoted in red in at the top of each panel.

872

873

874
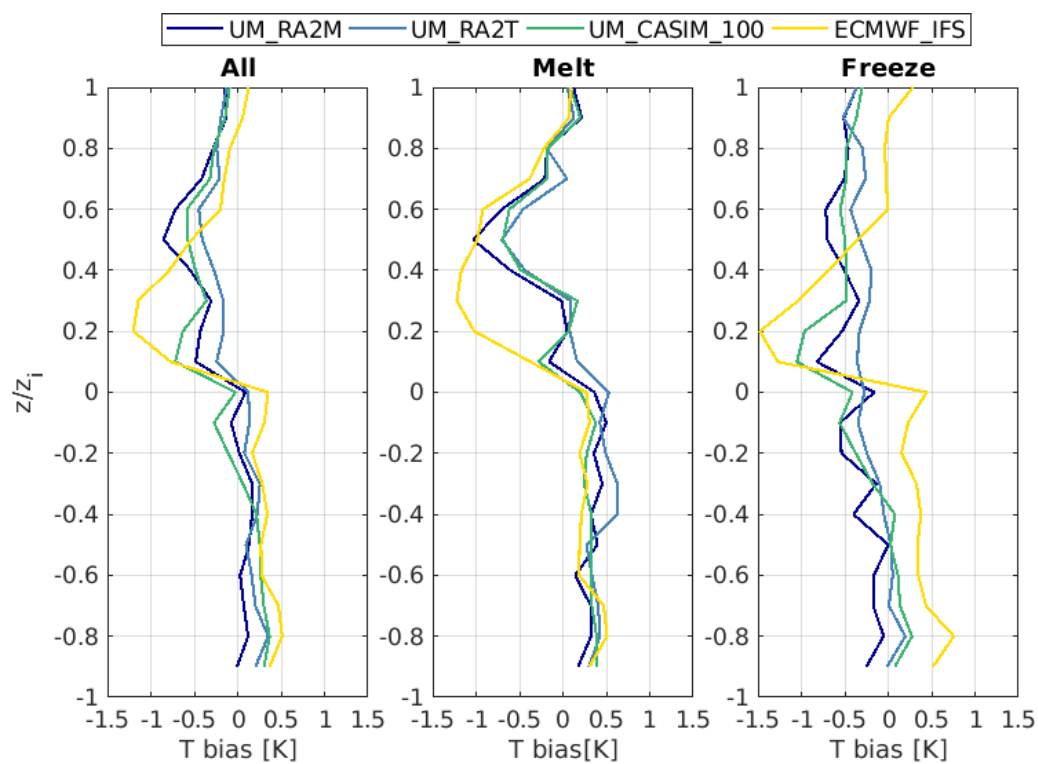
Figure 15: Scaled median model-observation T bias profiles for the full drift *(left)*, melt *(middle)*, and freeze *(right)* periods. Profiles are scaled such that –1 is the surface, 0 is the main inversion base, and 1 is 3 km.

875
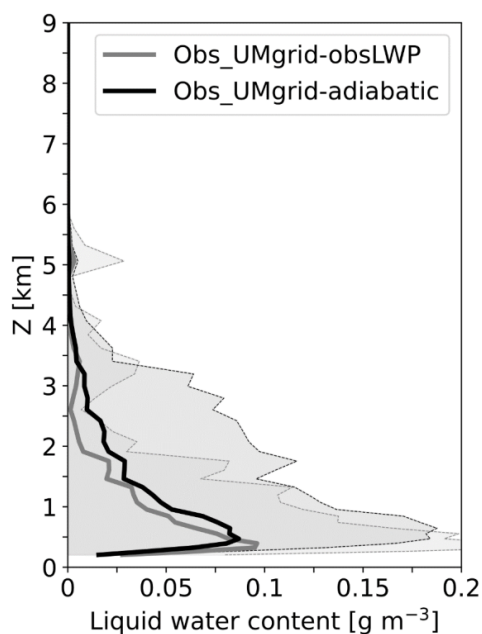
Figure A1: Comparison of mean LWC profiles calculated using an adiabatic assumption (black, ± one standard deviation shown in dark grey shading) and from HATPRO LWP measurements (grey, ± one standard deviation shown in light grey) without the Cloudnet offsetting procedure.

Data availability

877 UK contributions, as well as selected other data, are available within the MOCCHA (*Microbiology-Ocean-Cloud Coupling in*

878 *the High Arctic*) data collection in the Centre for Environmental Data Analysis (CEDA) archives (http://archive.ceda.ac.uk/).

879 Other cruise data are available from the Bolin Centre for Climate Research MOCCHA/AO2018 holdings

880 (http://bolin.su.se/data).

Author Contributions

882 GY led the model data analysis, aided by PF, RP, JD, and RF. JV led the measurement and Cloudnet data analysis, with

883 contributions from PA, IMB, MT, and EO. IMB, PA, MT, and JP performed the measurements during AO2018. IMB, MT, JD,

884 PF, RN, GY and JV all contributed to the study design. All authors contributed to the discussion of results and writing of the

885 manuscript.

Competing Interests

887 The authors declare that they have no conflict of interest.

Atmospheric
Chemistry
and Physics
Discussions

901 **References**

902 **Abdul-Razzak and Ghan**: A Parameterization of Aerosol Activation, Part II: Multiple Aerosol Types, *J. Geophys. Res.*, 105,
903 6837–6844, doi:10.1029/1999JD901161, 2000

904 **Acosta Navarro**, et al.: Link between autumnal Arctic sea ice and Northern Hemisphere winter forecast skill. *Geophysical*
905 *Research Letters*, 47, e2019GL086753. https://doi.org/10.1029/2019GL086753, 2020.

906 **Abel and Boutle**: An improved representation of the raindrop size distribution for single-moment microphysics schemes.
907 Q.J.R. Meteorol. Soc., 138: 2151-2162. doi:10.1002/qj.1949, 2012.

908 **Achtert**, P., O'Connor, E. J., Brooks, I. M., Sotiropoulou, G., Shupe, M. D., Pospichal, B., Brooks, B. J., and Tjernström, M.:
909 Properties of Arctic liquid and mixed-phase clouds from shipborne Cloudnet observations during ACSE 2014, Atmos. Chem.
910 Phys., https://doi.org/10.5194/acp-20-14983-2020 , 2020.

911 **Allen**, R. J., S. C. Sherwood, J. R. Norris, and C. S. Zender, 2012: Recent Northern Hemisphere tropical expansion primarily
912 driven by black carbon and tropospheric ozone. Nature, 485, 350–354.

913 **AMAP**, 2021. Arctic Climate Change Update 2021: Key Trends and Impacts. Summary for Policy-makers. Arctic Monitoring
914 and Assessment Programme (AMAP), Tromsø, Norway. 16 pp

915 **Bender**, F.A.-M., V. Ramanathan, and G. Tselioudis, 2012: Changes in extratropical storm track cloudiness 1983-2008:
916 Observational support for a poleward shift. *Clim. Dyn*., 38, 2037-2053, doi:10.1007/s00382-011-1065-6.

917 **Bigg**, E. K.: The formation of atmospheric ice crystals by the freezing of droplets, *Quarterly Journal of the Royal*
918 *Meteorological Society*, 79, 510–519, doi: 10.1002/qj.49707934207, 1953.

Atmospheric
Chemistry
and Physics
Discussions

919   **Birch**, C. E., Brooks, I. M., Tjernström, M., Milton, S. F., Earnshaw, P., Söderberg, S., and Persson, P. O. G. (2009), The
920   performance of a global and mesoscale model over the central Arctic Ocean during late summer, J. Geophys. Res., 114,
921   D13104, doi:10.1029/2008JD010790.

922   **Birch**, et al.: Modelling atmospheric structure, cloud and their response to CCN in the central Arctic: ASCOS case studies,
923   *Atmos. Chem. Phys.*, 12, 3419–3435, https://doi.org/10.5194/acp-12-3419-2012, 2012.

924   **Boeke and Taylor**: Evaluation of the Arctic surface radiation budget in CMIP5 models, *J. Geophys. Res. Atmos.*, 121, 8525–
925   8548, doi:10.1002/2016JD025099, 2016.

926   **Boucher**, et al.: Clouds and Aerosols, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group
927   I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner,
928   G. K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press,
929   Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/CBO9781107415324.016, 2013.

930   **Buizza**, R., J.-R. Bidlot, M. Janousek, S. Keeley, K. Mogensen & D. Richardson, 2017: New IFS cycle brings sea-ice coupling
931   and higher ocean resolution, *ECMWF Newsletter* No. 150, 14–17.

932   **Bush**, M., Allen, T., Bain, C., Boutle, I., Edwards, J., Finnenkoetter, A., Franklin, C., Hanley, K., Lean, H., Lock, A., Manners,
933   J., Mittermaier, M., Morcrette, C., North, R., Petch, J., Short, C., Vosper, S., Walters, D., Webster, S., Weeks, M., Wilkinson,
934   J., Wood, N., and Zerroukat, M.: The first Met Office Unified Model–JULES Regional Atmosphere and Land configuration,
935   RAL1, Geosci. Model Dev., 13, 1999–2029, https://doi.org/10.5194/gmd-13-1999-2020, 2020.

936   **Bourassa**, et al.: High-Latitude Ocean and Sea Ice Surface Fluxes: Challenges for Climate Research. *Bull. Amer. Meteor. Soc.*,
937   94, 403–423, https://doi.org/10.1175/BAMS-D-11-00244.1, 2013.

938   **Chylek**, P., T. J. Vogelsang, J. D. Klett, N. Hengartner, D. Higdon, G. Lesins, and M. K. Dubey, 2016: Indirect Aerosol Effect
939   Increases CMIP5 Models' Projected Arctic Warming. J. Climate, 29, 1417–1428, https://doi.org/10.1175/JCLI-D-15-0362.1.

940   **Cohen**, J., Screen, J., Furtado, J. et al. Recent Arctic amplification and extreme mid-latitude weather. Nature Geosci 7, 627–
941   637 (2014). https://doi.org/10.1038/ngeo2234

942   **Cooper**, W. A.: Ice Initiation in Natural Clouds, *Meteorological Monographs*, 21, 29–32, doi:10.1175/0065-9401-21.43.29,
943   1986.

944   **Cuxart**, J., Holtslag, A.A.M., Beare, R.J. *et al.* Single-Column Model Intercomparison for a Stably Stratified Atmospheric
945   Boundary Layer. *Boundary-Layer Meteorol* **118**, 273–303 (2006). https://doi.org/10.1007/s10546-005-3780-1

946   **de Boer**, G., Morrison, H., Shupe, M. D., and Hildner, R.: Evidence of liquid dependent ice nucleation in high-latitude
947   stratiform clouds from surface remote sensors, Geophys. Res. Lett., 38, L01803, doi:10.1029/2010GL046016, 2011.

948   **Dee**, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G.,
949   Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer,

950 A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally,
951 A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F.
952 (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc.,
953 137: 553-597. https://doi.org/10.1002/qj.828

954 **Ebert and Curry** (1993), An intermediate one-dimensional thermodynamic sea ice model for investigating ice-atmosphere
955 interactions, *J. Geophys. Res.*, 98(C6), 10085– 10109, doi:10.1029/93JC00656.

956 **Field**, et al.: Parametrization of ice-particle size distributions for mid-latitude stratiform cloud. *Quart. J. Royal Meterol. Soc.*,
957 131, pp 1997-2017, 2005.

958 **Field**, P. R., A. J. Heymsfield, and A. Bansemer, 2007: Snow Size Distribution Parameterization for Midlatitude and Tropical
959 Ice Clouds. J. Atmos. Sci., 64, 4346–4365, https://doi.org/10.1175/2007JAS2344.1.

960 **Fletcher**: The Physics of Rain Clouds. *Cambridge Univ Press*, Cambridge, UK, 1962.

961 **Forbes and Ahlgrimm,** 2014: On the Representation of High-Latitude Boundary Layer Mixed-Phase Cloud in the ECMWF
962 Global Model. *Mon. Wea. Rev.,* **142**, 3425–3445, https://doi.org/10.1175/MWR-D-13-00325.1

963 **Furtado**, K., P. R. Field, I. A. Boutle, C. J. Morcrette, and J. M. Wilkinson, 2016: A Physically Based Subgrid Parameterization
964 for the Production and Maintenance of Mixed-Phase Clouds in a General Circulation Model. J. Atmos. Sci., 73, 279–
965 291, https://doi.org/10.1175/JAS-D-15-0021.1.

966 **Gilbert**, et al.: Summertime cloud phase strongly influences surface melting on the Larsen C ice shelf, Antarctica. *Q J R*
967 *Meteorol Soc*.; 1– 16. https://doi.org/10.1002/qj.3753, 2020

968 **Grosvenor**, D. P., Field, P. R., Hill, A. A., and Shipway, B. J.: The relative importance of macrophysical and cloud albedo
969 changes for aerosol-induced radiative effects in closed-cell stratocumulus: insight from the modelling of a case study, Atmos.
970 Chem. Phys., 17, 5155–5183, https://doi.org/10.5194/acp-17-5155-2017, 2017

971 **Held and Soden**, 2006. Robust responses of the hydrological cycle to global warming. J. Climate 19: 5686–5699

972 **Hersbach**, et al.: The ERA5 Global Reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999-2049,
973 DOI:10.1002/qj.3803.

974 **Hill**, et al.: How sensitive are aerosol-precipitation interactions to the warm rain representation?, *J. Adv. Model. Earth Syst.*, 7,
975 987–1004, doi:10.1002/2014MS000422, 2015.

976 **Hines and Bromwich**: Simulation of Late Summer Arctic Clouds during ASCOS with Polar WRF. *Mon. Wea. Rev.,* **145**, 521–
977 541, https://doi.org/10.1175/MWR-D-16-0079.1, 2017.

978 **Hodson**, *et al.:* Identifying uncertainties in Arctic climate change projections. *Clim Dyn* 40, 2849–2865.
979 https://doi.org/10.1007/s00382-012-1512-z, 2013.

Atmospheric
Chemistry
and Physics
Discussions

980 **Hogan**, R. J., C. Jakob, and A. J. Illingworth, 2001: Comparison of ECMWF Winter-Season Cloud Fraction with Radar-
981 Derived Values. J. Appl. Meteor., 40, 513–525, https://doi.org/10.1175/1520-0450(2001)040<0513:COEWSC>2.0.CO;2.

982 **Illingworth**, et al.: Cloudnet, *Bull. Am. Meteorol. Soc.,* 88(6), 883–898, doi:10.1175/BAMS-88-6-883, 2007.

983 **Jakobson**, E., Vihma, T., Palo, T., Jakobson, L., Keernik, H., and Jaagus, J. (2012), Validation of atmospheric reanalyses over
984 the central Arctic Ocean, Geophys. Res. Lett., 39, L10802, doi:10.1029/2012GL051591.

985 **Jung**, T., and Coauthors, 2016: Advancing Polar Prediction Capabilities on Daily to Seasonal Time Scales. Bull. Amer. Meteor.
986 Soc., 97, 1631–1647, https://doi.org/10.1175/BAMS-D-14-00246.1.

987 **Karlsson and Svensson**: The simulation of Arctic clouds and their influence on the winter surface temperature in present-day
988 climate in the CMIP3 multi-model dataset. Clim Dyn 36, 623–635 (2011). https://doi.org/10.1007/s00382-010-0758-6

989 **Keeley and Mogensen**, 2018: Dynamic sea ice in the IFS, *ECMWF Newsletter* No. 156, 23—29.

990 **Kupiszewski**, et al.: Vertical profiling of aerosol particles and trace gases over the central Arctic Ocean during summer, *Atmos.*
991 *Chem. Phys.*, 13, 12405–12431, 2013.

992 **Lawrence**, H, Bormann, N, Sandu, I, Day, J, Farnan, J, Bauer, P. Use and impact of Arctic observations in the ECMWF
993 Numerical Weather Prediction system. *Q J R Meteorol Soc*. 2019; 145: 3432– 3454. https://doi.org/10.1002/qj.3628

994 **Mann**, G. W., Carslaw, K. S., Spracklen, D. V., Ridley, D. A., Manktelow, P. T., Chipperfield, M. P., Pickering, S. J., and
995 Johnson, C. E.: Description and evaluation of GLOMAP-mode: a modal global aerosol microphysics model for the UKCA
996 composition-climate model, Geosci. Model Dev., 3, 519–551, https://doi.org/10.5194/gmd-3-519-2010, 2010.

997 **Martin**, G. M., D. W. Johnson, and A. Spice, 1994: The Measurement and Parameterization of Effective Radius of Droplets
998 in   Warm   Stratocumulus   Clouds.   J.   Atmos.   Sci.,   51,   1823–1842,   https://doi.org/10.1175/1520-
999 0469(1994)051<1823:TMAPOE>2.0.CO;2.

1000 **Mauritsen**, T., G. Svensson, S. S. Zilitinkevich, I. Esau, L. Enger, and B. Grisogono, 2007: A Total Turbulent Energy Closure
1001 Model   for   Neutrally   and   Stably   Stratified   Atmospheric   Boundary   Layers.   *J.   Atmos.   Sci.*,   **64**,   4113–4126,
1002 https://doi.org/10.1175/2007JAS2294.1.

1003 **Mauritsen**,   et   al.:   An   Arctic   CCN-limited   cloud-aerosol   regime,   Atmos.   Chem.   Phys.,   11,   165–173,
1004 https://doi.org/10.5194/acp-11-165-2011, 2011.

1005 **Meyers**, et al.: New primary ice-nucleation parameterizations in an explicit cloud model. *J. Appl. Meteorol.* 31:708–721, 1992.

1006 **Miltenberger**, A. K., Field, P. R., Hill, A. A., Rosenberg, P., Shipway, B. J., Wilkinson, J. M., Scovell, R., and Blyth, A. M.:
1007 Aerosol–cloud interactions in mixed-phase convective clouds – Part 1: Aerosol perturbations, *Atmos. Chem. Phys.*, 18, 3119–
1008 3145, https://doi.org/10.5194/acp-18-3119-2018, 2018.

1009   **Morgenstern**, O., Braesicke, P., O'Connor, F. M., Bushell, A. C., Johnson, C. E., Osprey, S. M., and Pyle, J. A.: Evaluation of
1010   the new UKCA climate-composition model – Part 1: The stratosphere, Geosci. Model Dev., 2, 43–57,
1011   https://doi.org/10.5194/gmd-2-43-2009, 2009.

1012   **Naakka**, T., Nygård, T., & Vihma, T. (2018). Arctic Humidity Inversions: Climatology and Processes, *Journal of Climate*,
1013   *31*(10), 3765-3787. Retrieved Apr 13, 2021, from https://journals.ametsoc.org/view/journals/clim/31/10/jcli-d-17-0497.1.xml

1014   **Nygård**, T., Valkonen, T., and Vihma, T.: Characteristics of Arctic low-tropospheric humidity inversions based on radio
1015   soundings, Atmos. Chem. Phys., 14, 1959–1971, https://doi.org/10.5194/acp-14-1959-2014, 2014.

1016   **O'Connor**, F. M., Johnson, C. E., Morgenstern, O., Abraham, N. L., Braesicke, P., Dalvi, M., Folberth, G. A., Sanderson, M.
1017   G., Telford, P. J., Voulgarakis, A., Young, P. J., Zeng, G., Collins, W. J., and Pyle, J. A.: Evaluation of the new UKCA climate-
1018   composition model – Part 2: The Troposphere, Geosci. Model Dev., 7, 41–91, https://doi.org/10.5194/gmd-7-41-2014, 2014.

1019   **Rogers and Yau**. A short course in cloud physics. Pergamon Press, Oxford, 3rd edition, 1989.

1020   **Sandu**, I., Beljaars, A., Bechtold, P., Mauritsen, T., and Balsamo, G., Why is it so difficult to represent stably stratified
1021   conditions in numerical weather prediction (NWP) models?, *J. Adv. Model. Earth Syst.*, 5, 117– 133, doi:10.1002/jame.20013.

1022   **Sedlar**, et al.: On the Relationship between Thermodynamic Structure and Cloud Top, and Its Climate Significance in the
1023   Arctic. *J. Climate,* **25**, 2374–2393, https://doi.org/10.1175/JCLI-D-11-00186.1, 2012.

1024   **Seidel**, et al.: Climatology of the planetary boundary layer over the continental United States and Europe, *J. Geophys. Res.*,
1025   117, D17106, doi:10.1029/2012JD018143, 2012.

1026   **Serreze**, M. C. and Barry, R. G.: Processes and impacts of Arctic amplification: A research synthesis, Global Planet. Change,
1027   77, 85–96, doi:10.1016/j.gloplacha.2011.03.004, 2011.

1028   **Shupe** et al.: Arctic Mixed-Phase Cloud Properties Derived from Surface-Based Sensors at SHEBA. *J. Atmos. Sci.*, 63, 697–
1029   711, https://doi.org/10.1175/JAS3659.1, 2006.

1030   **Shupe**, M. D., Daniel, J. S., de Boer, G., Eloranta, E. W., Kollias, P., Long, C. N., Luke, E. P., Turner, D. D., and Verlinde, J.:
1031   A Focus On Mixed-Phase Clouds, Bulletin of the American Meteorological Society, 89, 1549–1562,
1032   doi:10.1175/2008BAMS2378.1, 2008.

1033   **Smith**: A scheme for predicting layer clouds and their water content in a general circulation model. *Q.J.R. Meteorol. Soc.*, 116:
1034   435-460. doi:10.1002/qj.49711649210, 1990.

1035   **Solomon**, A., Shupe, M. D., Persson, P. O. G., and Morrison, H.: Moisture and dynamical interactions maintaining decoupled
1036   Arctic mixed-phase stratocumulus in the presence of a humidity inversion, Atmos. Chem. Phys., 11, 10127–10148,
1037   https://doi.org/10.5194/acp-11-10127-2011, 2011.

1038   **Solomon**, A., Feingold, G., and Shupe, M. D.: The role of ice nuclei recycling in the maintenance of cloud ice in Arctic mixed-
1039   phase stratocumulus, Atmos. Chem. Phys., 15, 10631–10643, https://doi.org/10.5194/acp-15-10631-2015, 2015.

1040   **Sotiropoulou**, et al.: The thermodynamic structure of summer Arctic stratocumulus and the dynamic coupling to the surface,
1041   *Atmos. Chem. Phys*., 14, 12573–12592, https://doi.org/10.5194/acp-14-12573-2014, 2014.

1042   **Sotiropoulou**, et al.: Summer Arctic clouds in the ECMWF forecast model: an evaluation of cloud parametrization schemes.
1043   *Q.J.R. Meteorol. Soc.*, 142: 387-400. doi:10.1002/qj.2658. 2016

1044   **Spracklen**, D. V., Carslaw, K. S., Merikanto, J., Mann, G. W., Reddington, C. L., Pickering, S., Ogren, J. A., Andrews, E.,
1045   Baltensperger, U., Weingartner, E., Boy, M., Kulmala, M., Laakso, L., Lihavainen, H., Kivekäs, N., Komppula, M.,
1046   Mihalopoulos, N., Kouvarakis, G., Jennings, S. G., O'Dowd, C., Birmili, W., Wiedensohler, A., Weller, R., Gras, J., Laj, P.,
1047   Sellegri, K., Bonn, B., Krejci, R., Laaksonen, A., Hamed, A., Minikin, A., Harrison, R. M., Talbot, R., and Sun, J.: Explaining
1048   global surface aerosol number concentrations in terms of primary emissions and particle formation, Atmos. Chem. Phys., 10,
1049   4775–4793, https://doi.org/10.5194/acp-10-4775-2010, 2010.

1050   **Stevens**, R. G., Loewe, K., Dearden, C., Dimitrelos, A., Possner, A., Eirund, G. K., Raatikainen, T., Hill, A. A., Shipway, B.
1051   J., Wilkinson, J., Romakkaniemi, S., Tonttila, J., Laaksonen, A., Korhonen, H., Connolly, P., Lohmann, U., Hoose, C., Ekman,
1052   A. M. L., Carslaw, K. S., and Field, P. R.: A model intercomparison of CCN-limited tenuous clouds in the high Arctic, *Atmos.*
1053   *Chem. Phys.*, 18, 11041–11071, https://doi.org/10.5194/acp-18-11041-2018, 2018.

1054   **Sundqvist**: A parameterization scheme for non-convective condensation including prediction of cloud water content. *Quart.*
1055   *J. R. Met. Soc.* 104, pp. 677490, 1978.

1056   **Tiedtke**: Representation of Clouds in Large-Scale Models. *Mon. Wea. Rev.,* **121**, 3040–3061, https://doi.org/10.1175/1520-
1057   0493(1993)121<3040:ROCILS>2.0.CO;2, 1993.

1058   **Tjernström** et al.: The Arctic Summer Cloud Ocean Study (ASCOS): overview and experimental design, *Atmos. Chem. Phys.*,
1059   14, 2823–2869, https://doi.org/10.5194/acp-14-2823-2014, 2014.

1060   **Tjernström**, M, Svensson, G, Magnusson, L, et al. Central Arctic weather forecasting: Confronting the ECMWF IFS with
1061   observations from the Arctic Ocean 2018 expedition. Q J R Meteorol Soc. 2021; 1– 22. https://doi.org/10.1002/qj.3971

1062   **Vignesh** et al.: Assessment of CMIP6 cloud fraction and comparison with satellite observations. Earth and Space Science, 7,
1063   e2019EA000975. https://doi.org/10.1029/2019EA000975, 2020.

1064   **Vihma** et al.: Advances in understanding and parameterization of small-scale physical processes in the marine Arctic climate
1065   system: a review, Atmos. Chem. Phys., 14, 9403–9450, doi:10.5194/acp-14-9403-2014, 2014.

1066   **Vüllers**, J., Achtert, P., Brooks, I. M., Tjernström, M., Prytherch, J., Burzik, A., and Neely III, R.: Meteorological and cloud
1067   conditions during the Arctic Ocean 2018 expedition, Atmos. Chem. Phys., 21, 289–314, https://doi.org/10.5194/acp-21-289-
1068   2021, 2021.

1069   **Walters**, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S., Wells, H., Williams, K., Wood, N., Allen, T., Bushell,
1070   A., Copsey, D., Earnshaw, P., Edwards, J., Gross, M., Hardiman, S., Harris, C., Heming, J., Klingaman, N., Levine, R.,
1071   Manners, J., Martin, G., Milton, S., Mittermaier, M., Morcrette, C., Riddick, T., Roberts, M., Sanchez, C., Selwood, P., Stirling,
1072   A., Smith, C., Suri, D., Tennant, W., Vidale, P. L., Wilkinson, J., Willett, M., Woolnough, S., and Xavier, P.: The Met Office

Atmospheric
Chemistry
and Physics
Discussions
Open Access

EGU

1073   Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations, Geosci. Model Dev., 10, 1487–

1074   1520, https://doi.org/10.5194/gmd-10-1487-2017, 2017.

1075   **Walters** et al.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, Geosci.

1076   Model Dev., 12, 1909–1963, https://doi.org/10.5194/gmd-12-1909-2019, 2019.

1077   **Wesslén**, C., Tjernström, M., Bromwich, D. H., de Boer, G., Ekman, A. M. L., Bai, L.-S., and Wang, S.-H.: The Arctic summer

1078   atmosphere: an evaluation of reanalyses using ASCOS data, Atmos. Chem. Phys., 14, 2605–2624, https://doi.org/10.5194/acp-

1079   14-2605-2014, 2014.

1080   **Wilson and Ballard**: A microphysically based precipitation scheme for the UK meteorological office unified model. Q.J.R.

1081   Meteorol. Soc., 125: 1607-1636. doi:10.1002/qj.49712555707, 1999.

1082   **Wilson**, D.R., Bushell, A.C., Kerr-Munslow, A.M., Price, J.D. and Morcrette, C.J. (2008), PC2: A prognostic cloud fraction

1083   and condensation scheme. I: Scheme description. *Q.J.R. Meteorol. Soc.*, 134: 2093-2107. doi:10.1002/qj.333

1084   **Wood and Field**: Relationships between Total Water, Condensed Water and Cloud Fraction in Stratiform Clouds Examined

1085   Using Aircraft Data, *J. Atmos. Sci*., 57, 1888–1905, 2000.

1086   **Varma**, V., Morgenstern, O., Furtado, K., Field, P., and Williams, J.: Introducing Ice Nucleating Particles functionality into

1087   the Unified Model and its impact on the Southern Ocean short-wave radiation biases, Atmos. Chem. Phys. Discuss. [preprint],

1088   https://doi.org/10.5194/acp-2021-438, in review, 2021.

1089   **Vavrus**, S., D. Waliser, A. Schweiger, and J. Francis, 2009: Simulations of 20th and 21st century Arctic cloud amount in the

1090   global climate models assessed in the IPCC AR4. *Clim. Dyn.*, 33, 1099–1115

1091   **Young**, G., Connolly, P. J., Jones, H. M., and Choularton, T. W.: Microphysical sensitivity of coupled springtime Arctic

1092   stratocumulus to modelled primary ice over the ice pack, marginal ice, and ocean, *Atmos. Chem. Phys.*, 17, 4209–4227,

1093   https://doi.org/10.5194/acp-17-4209-2017, 2017.

1094   **Zuidema**, P., et al., 2005: An Arctic springtime mixed-phase cloudy boundary layer observed during SHEBA. *J. Atmos. Sci.*,

1095   62, 160–176.