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

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)

S1 Cloudnet

S1.1 Cloud fractions

Cloudnet produces a cloud fraction variable, C_V, in each model output file, which represents the Cloudnet cloud fraction calculated from observational data (from radar, lidar etc.) combined with the temperature and humidity profiles defined by the filename. For example, C_V in the UM_RA2M output corresponds to model temperature and humidity profiles combined with retrieved cloud properties from the remote sensing instruments on board *Oden* to produce a defined cloud fraction. These C_V variables from the observation, UM_RA2M, and ECMWF_IFS Cloudnet output files are shown in **Fig. S1**. *Obs_Cv* represents the measured/retrieved temperature profiles with reference to the radar vertical grid. **Figure S1** therefore demonstrates that the chosen grid on which the C_V data are shown have little impact on the mean profile.

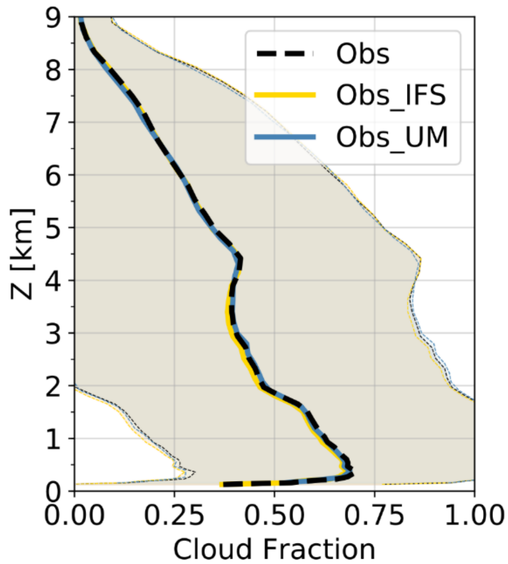

Figure S1: Comparison of cloud fraction (C_V) on native radar (Obs), IFS (Obs_IFS), and UM LAM (Obs_UM) vertical grid.

Poor comparisons between modelled and observed cloud fractions is a perennial problem in climate science, and our results indicate that the large-scale cloud scheme used to represent sub-grid-scale variability in the RH field may be responsible for producing particularly poor comparisons with observations. Cloudnet represents sub-grid-scale variability in observed cloud fractions by assigning a value of 0 or 1 at each vertical point, dependent on whether there is any form of cloud water present (including frozen precipitation), before data are averaged from the raw sample frequency to the model grid. This assignment is not necessarily equivalent to a model's bulk cloud fraction by volume, C_V, given the parameterisation of sub-grid-scale processes in models is imposed after the calculation of advected parameters. This discrepancy is not a new finding; **Illingworth et al., (2007)** noted the UM – both global and mesoscale variants using the **Smith (1990)** large-scale cloud scheme – had difficulty with simulating completely cloudy grid boxes. Similarly, **Hogan et al. (2001)** have previously shown that the ECMWF IFS underpredicts the fraction of observed cloud below 7 km.

Cloud fractions are diagnostic for UM_RA2M and UM_CASIM-100 (following **Smith, 1990**); prognostic for ECMWF_IFS and UM_RA2T (using PC2 scheme; **Wilson et al., 2008**). Both diagnostic and prognostic approaches use the instantaneous condensation assumption applied to a PDF of moisture and temperature (**Bush et al., 2020**). In the **Smith (1990)** scheme, the liquid and ice cloud fractions, C_{liq} and C_{ice}, are diagnosed from prognostic grid-box mean liquid and ice mass mixing ratios, q'_{liq} and q'_{ice} (**Wilson et al., 2008**; **Bush et al., 2020**), which are then combined assuming minimum overlap to compute C_V. The C_{liq} would be 0.5 when the grid-box mean total specific humidity is at saturation for the given temperature, since the parameterisation PDF of sub-grid-scale variability in RH and temperature is symmetric (**Wilson et al., 2008**). The **Smith (1990)** scheme was designed this way to keep the RH and bulk cloud fraction at realistic values (less than 1) over large grid-boxes. In our UM simulations, this cloud fraction is supplemented by an empirical adjustment based on aircraft observations (**Wood and Field, 2000**) which affects the rate at which cloud fraction increases once RH_{crit} is reached, increasing C_V up to 0.7 at 100%

RH; however, this adjustment is still insufficient to attain the cloud fractions approximately equal to 1 obtained from our observations.

Both the PC2 scheme in UM_RA2T and the cloud scheme in ECMWF_IFS are based on **Tiedtke (1993)**; both use prognostic cloud fraction and condensate variables, with the former calculated directly from condensate sources/sinks rather than being linked to the grid-box mean liquid water mixing ratio (as in the **Smith 1990** scheme; **Forbes and Ahlgrimm, 2014**). In the PC2 scheme, C_{liq} is not diagnosed from q'_{liq}; therefore, autoconversion does not affect C_{liq}, but does alter q'_{liq}, allowing thin clouds with low q'_{liq} to maintain a high C_{liq} (**Wilson et al., 2008**; **Bush et al., 2020**). Given the results described here, it appears that this functionality is critical to replicating cloud fractions comparable to those calculated from observations using Cloudnet. UM_RA2T also has the extra source of sub-grid turbulent production of mixed-phase cloud; in a test with this option switched off it was found that this process does not account for the improved cloud fraction comparison with observations (**Fig. S2**).

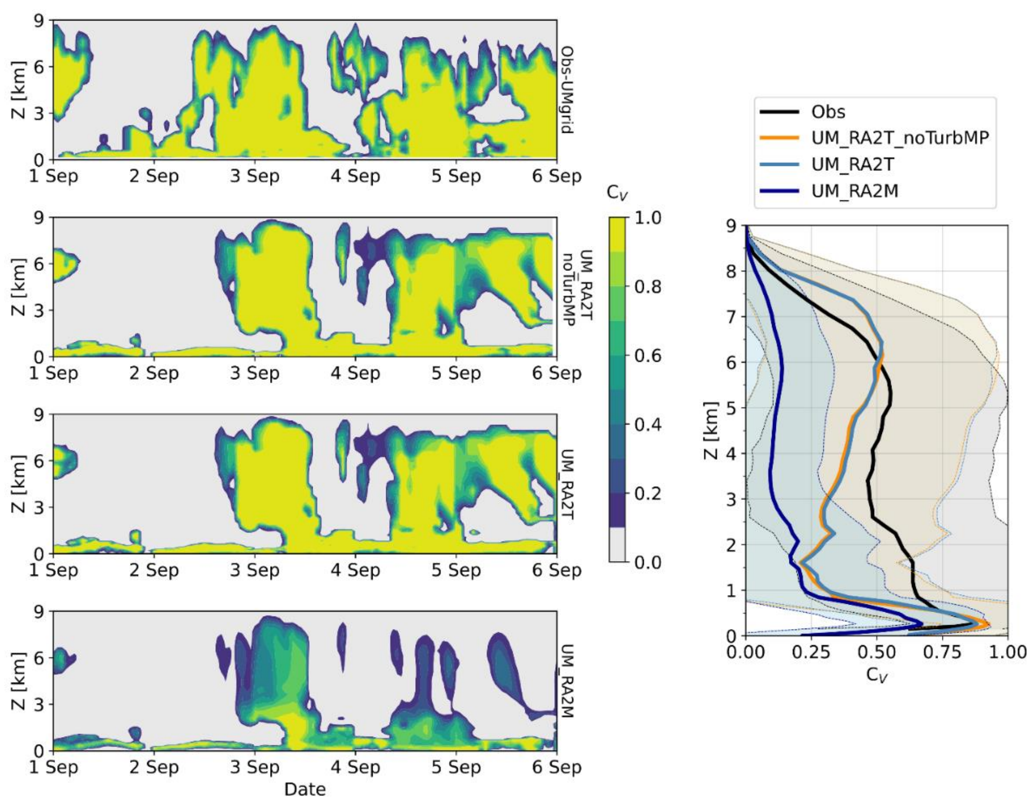

Figure S2: C_V (**a**) observed and modelled by (**b**) UM_RA2T without sub-grid turbulent production of mixed-phase cloud switched on (UM_RA2T_noTurbMP), (**c**) UM_RA2T, and (**d**) UM_RA2M for a subset of the drift period (1 Sep to 6 Sep), illustrating that the inclusion of this sub-grid process does not account for the high cloud fractions modelled by the UM_RA2T case.

With 3 separate cloud fractions instead of one total fraction (as represented in **Smith, 1990**), the PC2 scheme can represent overlap between the liquid and ice fractions, i.e., a mixed-phase cloud fraction (**Wilson et al., 2008**). This ability is likely important in reproducing the common mixed-phase clouds in the Arctic. The liquid cloud fraction represented by PC2 is equivalent to the total cloud fraction diagnosed by **Smith (1990)**, whereas the ice cloud fraction is calculated from the prognostic q_{ice}. The total cloud fraction is then the volume of a grid-box containing cloud, calculated assuming minimum overlap, between the liquid and ice cloud fractions (**Wilson et al., 2008**). As such, PC2 can represent a wide range of IWC for the same cloud fraction, whereas this relationship is fixed for a given temperature with the **Smith (1990)** scheme.

It is important to note that any frozen precipitation measured below an upper cloud layer will also be classed as cloud by Cloudnet. Some of the discrepancy in cloud fraction may then be due to observed precipitating clouds masking several cloud layers, while this layering may be captured by our models. However, one must note that the **Wilson and Ballard (1999)** microphysics scheme makes this same assumption regarding frozen precipitation. Alternatively, little-to-no precipitation between layers in our models would negatively affect this cloud fraction comparison. By treating all cloud ice the same, it may be difficult to distinguish between multi-layered clouds using Cloudnet if any are precipitating.

In summary, our data highlight the difficulty of diagnostic cloud fractions, such as that computed by the **Smith (1990)** scheme, to produce fully cloudy grid boxes on fine grid scales and indicate that models represent cloud fractions in different ways; therefore, cloud water content analyses should be conducted in addition to cloud fraction comparisons when studying Arctic clouds to provide a more robust model-observation comparison than using cloud fractions alone. **Figure S3** demonstrates how using a TWC threshold to define in- and out-of-cloud regions can give a more consistent comparison between the observations and models.

S1.2 Averaging observations to model grid

Liquid and ice water contents are calculated at Stage 2a of Cloudnet (**Illingworth et al., 2007**) using the various measured inputs described in **Sect. 2.2** of this study and referencing the radar height grid and time resolution. For comparison with numerical models, Cloudnet includes several additional functions in Stage 2b to average these observational data on to the corresponding model grids in a consistent manner.

Observational data are split into each model grid box then statistics are calculated, e.g., grid-box mean liquid water content. To ensure that there are enough data present in each box, a quality factor is applied: this factor is defined by default to be related to 90% of the grid box size, designed at Cloudnet's creation to ensure that there were enough observational data within each > 10 km model grid box for meaningful statistical comparisons. However, we found that this high value for the quality-control factor was too efficient in filtering out data with our higher spatial resolution (1.5 km) grid boxes.

This factor was reduced to 10% or 30% respectively for the IFS and UM grids utilised in this study to reduce the number of profiles required for meaningful statistics within each 4D box. Code failures relating to too few data restricted the UM quality factor being reduced further than 30%, but this is unsurprising given each box is 1.5 × 1.5 km in this study.

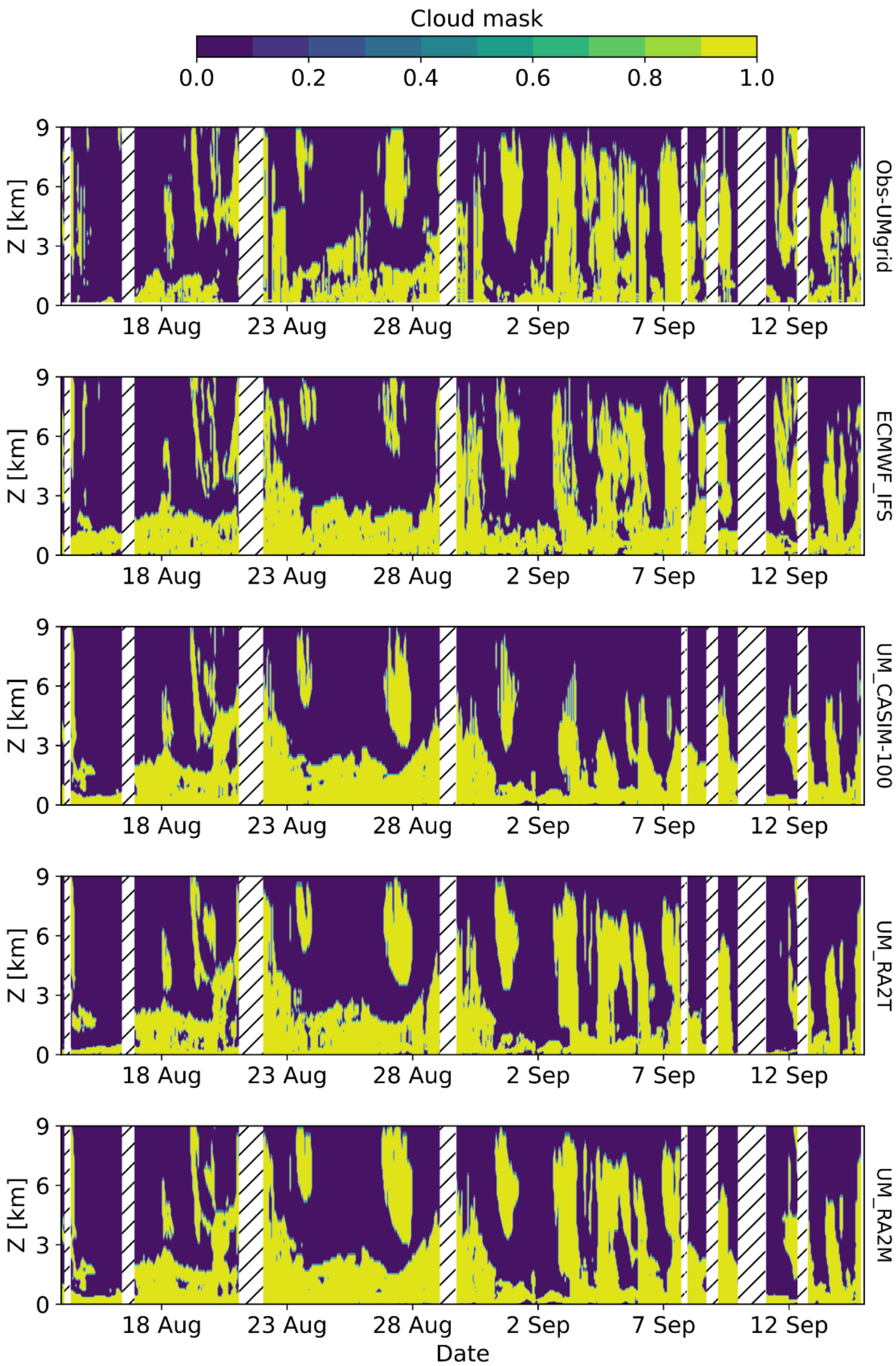

Figure S3: Comparison of cloud masks built using the in-cloud TWC threshold described in **Sect. 2.4**. TWC is calculated using all LWC and IWC data, then thresholding is applied.

S2 Model surface albedo

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

The surface albedo parameterisation used within the global and regional UM is dependent on ice surface temperature, where the relationship itself is unchanged from **Birch et al. (2009)**. Both UM_RA2M and UM_RA2T use the default Regional Atmosphere surface albedo thresholds, giving a 50% albedo at 0 °C which increases to 80% at –10 °C. **Gilbert et al. (2020)** tested both configurations for polar cloud modelling over the Antarctic Peninsula, finding that the surface albedo was modelled to within 2% of observed values.

For UM_CASIM-100, we adapted the warm ice temperature albedo of the LAM to 72% (at 0 °C), with 80% albedo achieved at –2 °C, to match the parameterisation limits currently used in the JULES (*Joint UK Land Environment Simulator*) surface scheme of the Global Atmosphere 6.0 global model (under the assumption that snow is present on the sea ice surface). For the drift period, we know that snow was indeed present on the surface from first-hand knowledge and surface imagery.

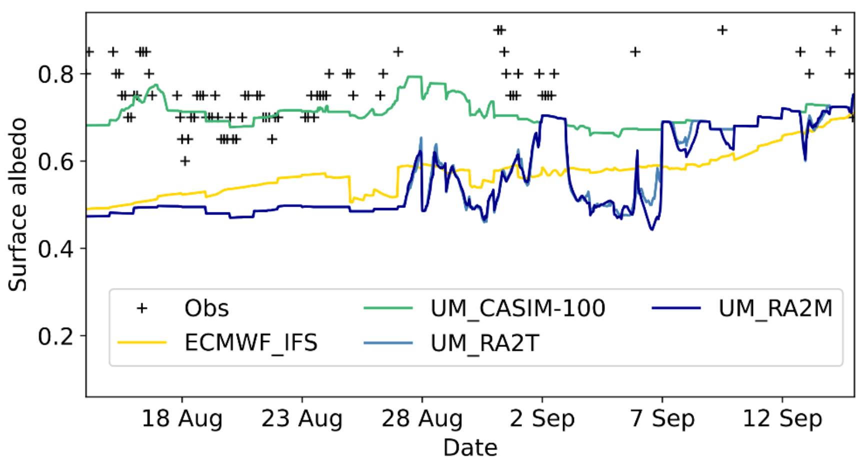

Figure S4: Surface albedo estimated from surface images of ice cover taken from the ship (**Obs**, black crosses) and diagnosed by the models.

Figure S4 shows model surface albedo as a function of time throughout the drift period, with estimations from ship-based observations shown as black crosses. The largest discrepancy between the models is during the melt period of the drift (before 28 Aug): UM_CASIM-100 performs well with comparison to our estimations, however the other three simulations underestimate by approximately 20%. Agreement between the models improves during the freeze when the surface temperature begins to fall; however, our few observational data points during this period suggest that the models are still underestimating albedo by approximately 10-15%.

While this comparison suggests that the models are performing poorly with regards to surface reflectivity, one must note that the models are representing the albedo of a 1.5/9 km grid box (UM/IFS, respectively) while the observed estimates are taken from the area immediately surrounding the ship. Therefore, any reduction in model albedo due to e.g., melt ponds or leads would not be accounted for in our observational estimates.

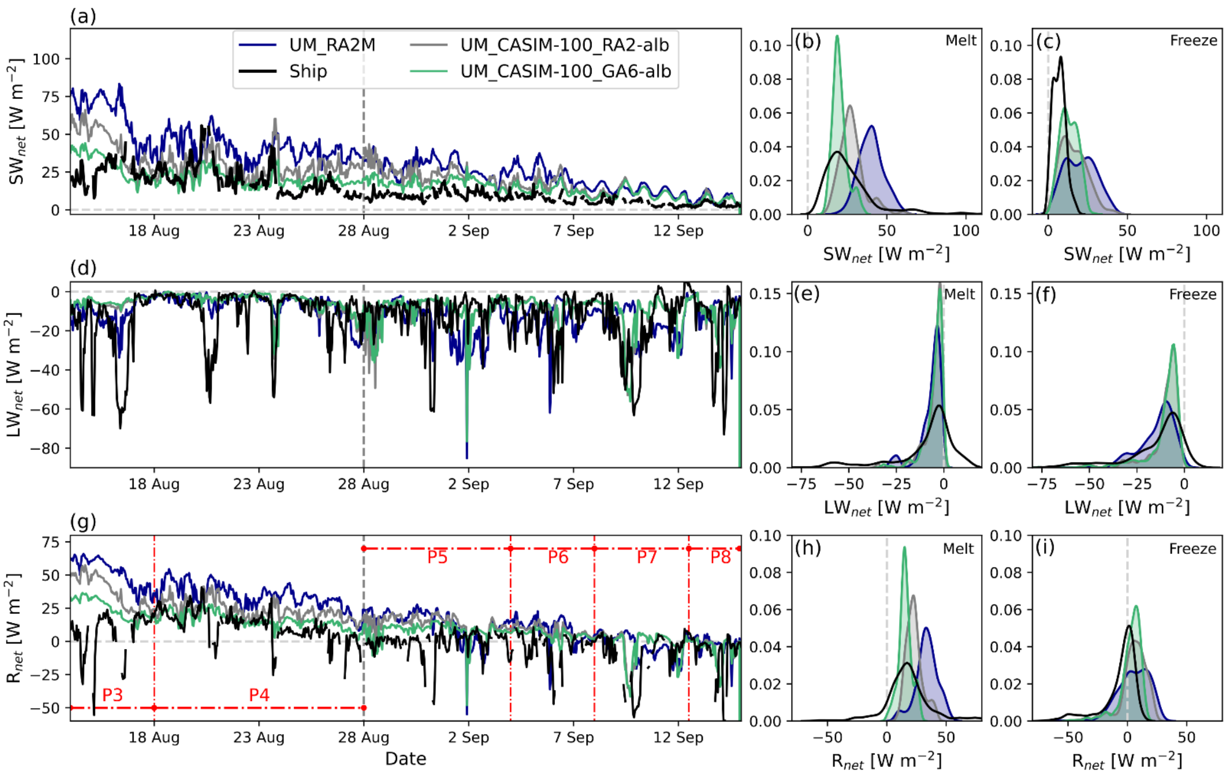

Figure S5: SW_{net}, LW_{net}, and R_{net} simulated by UM_CASIM-100 with albedo options for the Regional Atmosphere version 2 (UM_CASIM-100_RA2-alb; grey), updated Global Atmosphere version 6.0 (UM_CASIM-100_GA6-alb; green) used in the main body of this study, and UM_RA2M (dark blue) for reference. Hourly-averaged measurements on board the ship (black) shown for comparison. LHS: timeseries; RHS: PDFs. PDFs are split between melting and re-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).

Figure S5 shows the surface radiative balance modelled in UM_CASIM-100 (as **Fig. 2**, here labelled UM_CASIM-100_GA6-alb) and UM_CASIM-100 with the default Regional Atmosphere limits for the surface albedo parameterisations used (labelled UM_CASIM-100_RA2-alb, as used in UM_RA2M and UM_RA2T). **Figure S5** therefore shows that the cloud physics representation of UM_CASIM-100 does still improve radiative interactions, with comparison with our observations, over UM_RA2M (and UM_RA2T, not shown). Thus, the surface albedo updates are not the sole cause of its improved performance over the operational UM schemes; however, the combination of the updated surface albedo (to represent snow on sea ice) and improved cloud microphysical representation (from the CASIM scheme) yields the best UM comparison with observations (as presented in the main body of this study).

S3 UK Chemistry and Aerosol (UKCA) model

UKCA simulates gas and aerosol chemistry and transport in the atmosphere using the GLObal Model of Aerosol Processes (GLOMAP-Mode, **Mann et al., 2010**) and an atmospheric chemistry scheme, with an additional boundary layer nucleation scheme used to simulate gas-to-particle conversion of sulphuric acid to sulphate aerosol (**Spracklen et al., 2010**). To generate the aerosol input files for CASIM, UKCA was one-way coupled to the UM at version 11.2 using the Global Atmosphere 7.1 dynamical core (**Walters et al., 2019**). Daily averaged soluble accumulation- and coarse-mode aerosol number and mass concentrations calculated from UKCA grid points north of 88.125 °N (**Fig. S2**) were used as input to the UM.

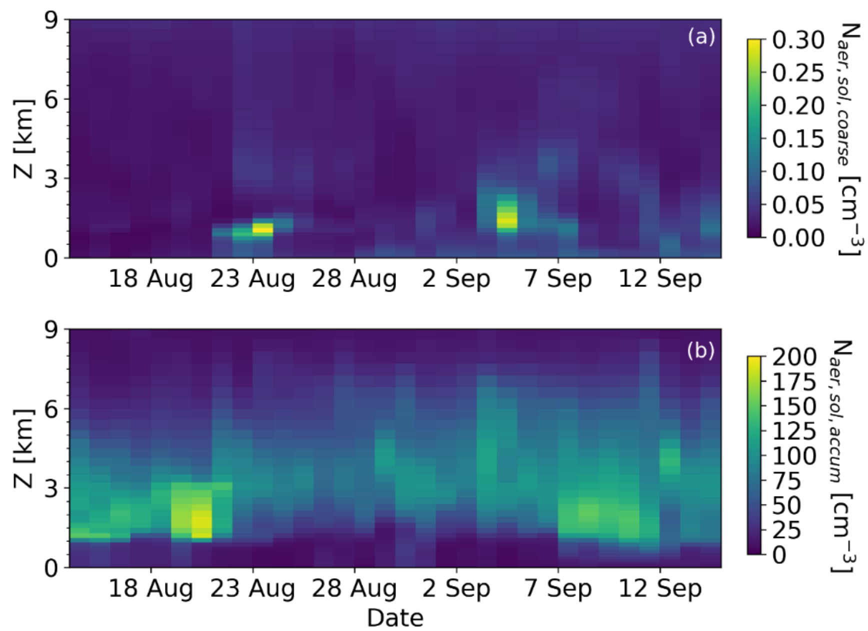

Figure S6: Soluble (**a**) coarse- and (**b**) accumulation-mode aerosol number concentrations simulated by UKCA and used as input for the CASIM scheme in the UM_CASIM-AeroProf simulation. Aerosol profiles are kept constant over each daily forecast, with no aerosol processing by cloud.

S4 Periods of consistent meteorology

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

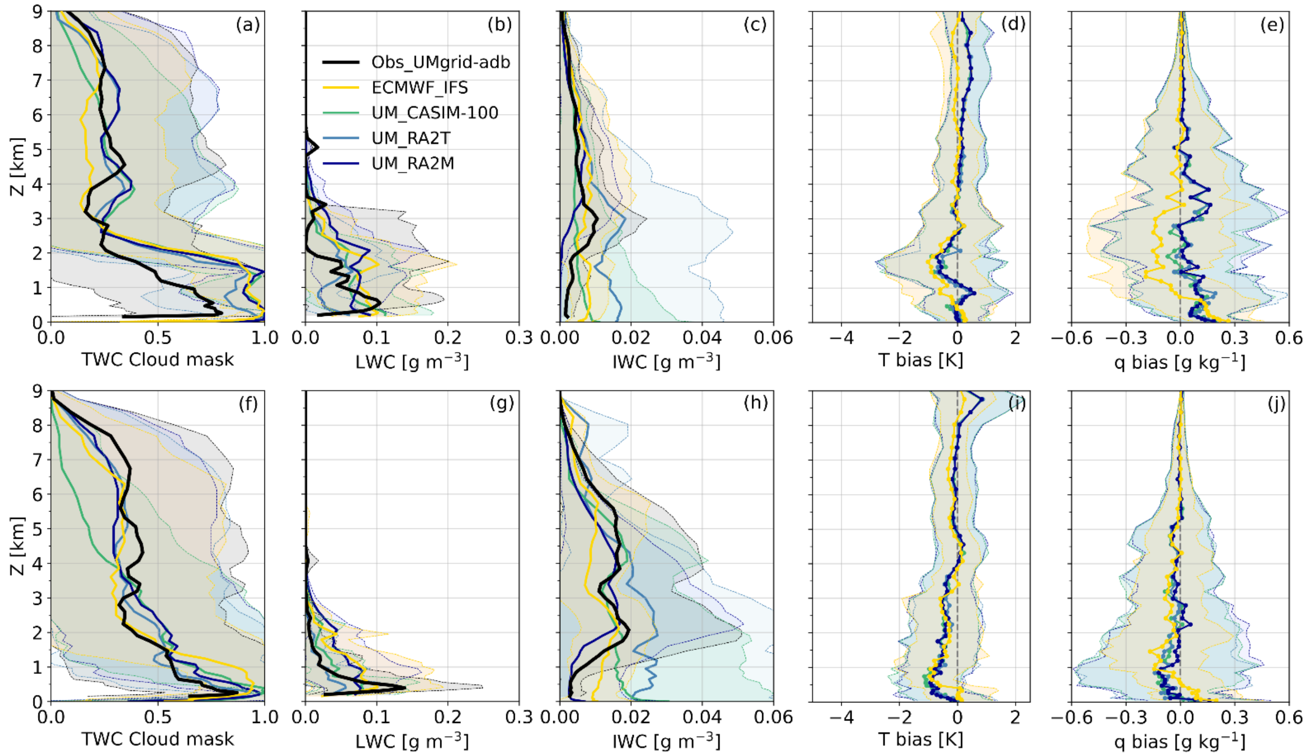

Figure S7: (As **Fig. 13)** Comparison of mean cloud mask, LWC, and IWC profiles with median biases in T and q with respect to radiosondes for period 4 (**a—e**, *top row*) and period 5 (**f—j**, *bottom row*). Again, observed LWC calculated assuming adiabatic conditions using Cloudnet. ± one standard deviation shown in shading to illustrate variability.

Figure S7 shows comparisons of the TWC cloud mask, in-cloud LWC and IWC, and associated T and q biases over periods 4 and 5. Both periods 4 and 5 support the findings of periods 3 and 6. During period 4, the models overpredict cloud occurrence below 2 km similar to period 3; however, both the LWC and IWC are in better agreement with observations during period 4. Similarly, all simulations agree better with observed cloud occurrence during period 5 (consistent with our result for period 6), and both the LWC and IWC are again in reasonable agreement during this time window. Consequently, the model thermodynamic biases with respect to radiosonde measurements are weaker (though still present) during periods 4 and 5 than during periods 3 and 6.

S5 Primary ice nucleation parameterisation

UM_RA2M and UM_RA2T use the **Fletcher (1962)** parameterisation for primary ice formation, while ECMWF_IFS uses **Meyers et al. (1992)** and UM_CASIM-100 uses **Cooper (1986)**. Each of these parameterisations is inherently temperature-dependent, with **Meyers et al., (1992)** producing the largest ice number concentration, and **Fletcher (1962)** producing the smallest, at e.g., –10 °C. To test whether the method of parameterising primary ice itself has any effect on these biases, we trialled the use of each of the **Fletcher (1962)**, **Cooper (1986)**, and **Meyers et al. (1992)** parameterisations within the CASIM framework; however, we found little difference in our tropospheric ice due to the different parameterisation methods (**Fig. S7**).

Changing the primary ice parameterisation alters biases slightly within the lowest 3 km of the domain, with a maximum difference of 0.06 g kg^{-1} between the UM_CASIM-100_Cooper and UM_CASIM-100_Meyers median *q* biases at 1.3 km. Differences shown here are much smaller than the more significant UM configurations changes/IFS comparison shown in **Figs. 9, 11, 13**.

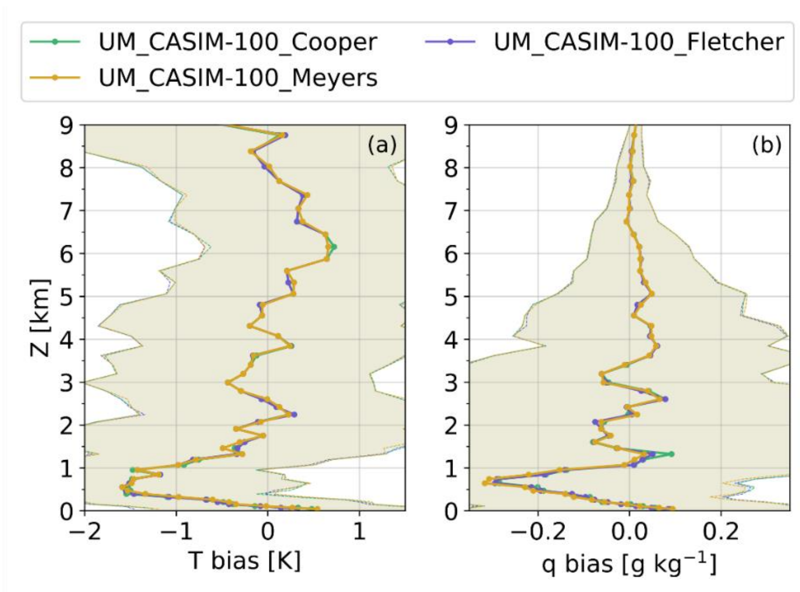

Figure S8: Comparison of *T* and *q* biases (with respect to radiosonde measurements) of CASIM-100 runs with different primary ice production parameterisations imposed (green: **Cooper, 1986**; purple: **Fletcher, 1962**; gold: **Meyers et al., 1992**), over the drift subset of 31 Aug to 5 Sep. Here, UM_CASIM-100_Cooper is equivalent to UM_CASIM-100 data shown in the main body of the paper.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

Walters et al.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, Geosci. Model Dev., 12, 1909–1963, https://doi.org/10.5194/gmd-12-1909-2019, 2019.

Wilson and Ballard: A microphysically based precipitation scheme for the UK meteorological office unified model. Q.J.R. Meteorol. Soc., 125: 1607-1636. doi:10.1002/qj.49712555707, 1999.

Wilson, D.R., Bushell, A.C., Kerr-Munslow, A.M., Price, J.D. and Morcrette, C.J. (2008), PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. *Q.J.R. Meteorol. Soc.*, 134: 2093-2107. doi:10.1002/qj.333

Wood and Field: Relationships between Total Water, Condensed Water and Cloud Fraction in Stratiform Clouds Examined Using Aircraft Data, *J. Atmos. Sci*., 57, 1888–1905, 2000.