Supplement of

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

Gillian Young McCusker et al.

Correspondence to: Gillian Young McCusker (g.y.mccusker@leeds.ac.uk) and Ian M. Brooks (i.m.brooks@leeds.ac.uk)

1   Supporting Information

2   **S1   Cloudnet**

3   **S1.1   Cloud fractions**

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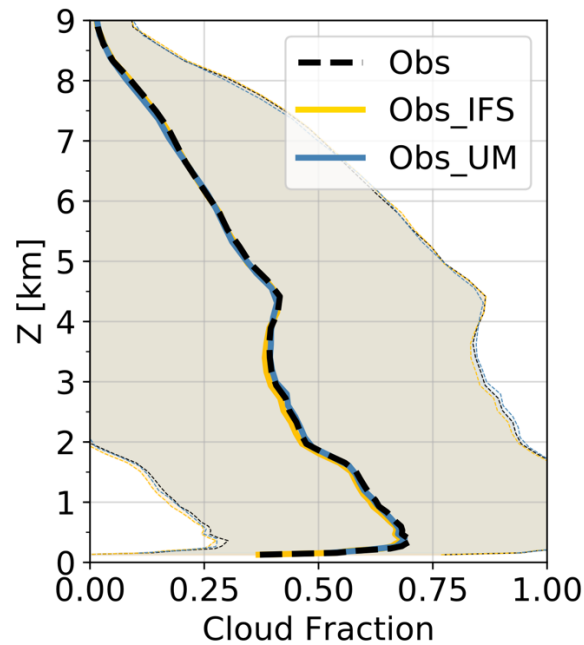
Figure S1: Comparison of cloud fraction (C_V) on native radar (Obs), IFS (Obs_IFS), and UM LAM (Obs_UM) vertical grid.

11   Poor comparisons between modelled and observed cloud fractions is a perennial problem in climate science, and our results
12   indicate that the large-scale cloud scheme used to represent sub-grid-scale variability in the RH field may be responsible for
13   producing particularly poor comparisons with observations. Cloudnet represents sub-grid-scale variability in observed cloud
14   fractions by assigning a value of 0 or 1 at each vertical point, dependent on whether there is any form of cloud water present
15   (including frozen precipitation), before data are averaged from the raw sample frequency to the model grid. This assignment is
16   not necessarily equivalent to a model's bulk cloud fraction by volume, C_V, given the parameterisation of sub-grid-scale

17 processes in models is imposed after the calculation of advected parameters. This discrepancy is not a new finding; **Illingworth**
18 **et al., (2007)** noted the UM – both global and mesoscale variants using the **Smith (1990)** large-scale cloud scheme – had
19 difficulty with simulating completely cloudy grid boxes. Similarly, **Hogan et al. (2001)** have previously shown that the
20 ECMWF IFS underpredicts the fraction of observed cloud below 7 km.

21 Cloud fractions are diagnostic for UM_RA2M and UM_CASIM-100 (following **Smith, 1990**); prognostic for ECMWF_IFS
22 and UM_RA2T (using PC2 scheme; **Wilson et al., 2008**). Both diagnostic and prognostic approaches use the instantaneous
23 condensation assumption applied to a PDF of moisture and temperature (**Bush et al., 2020**). In the **Smith (1990)** scheme, the
24 liquid and ice cloud fractions, C_{liq} and C_{ice}, are diagnosed from prognostic grid-box mean liquid and ice mass mixing ratios,
25 q'_{liq} and q'_{ice} (**Wilson et al., 2008**; **Bush et al., 2020**), which are then combined assuming minimum overlap to compute C_V.
26 The C_{liq} would be 0.5 when the grid-box mean total specific humidity is at saturation for the given temperature, since the
27 parameterisation PDF of sub-grid-scale variability in RH and temperature is symmetric (**Wilson et al., 2008**). The **Smith (1990)**
28 scheme was designed this way to keep the RH and bulk cloud fraction at realistic values (less than 1) over large grid-boxes. In
29 our UM simulations, this cloud fraction is supplemented by an empirical adjustment based on aircraft observations (**Wood and**
30 **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%
31 RH; however, this adjustment is still insufficient to attain the cloud fractions approximately equal to 1 obtained from our
32 observations.

33 Both the PC2 scheme in UM_RA2T and the cloud scheme in ECMWF_IFS are based on **Tiedtke (1993)**; both use prognostic
34 cloud fraction and condensate variables, with the former calculated directly from condensate sources/sinks rather than being
35 linked to the grid-box mean liquid water mixing ratio (as in the **Smith 1990** scheme; **Forbes and Ahlgrimm, 2014**). In the
36 PC2 scheme, C_{liq} is not diagnosed from q'_{liq}; therefore, autoconversion does not affect C_{liq}, but does alter q'_{liq}, allowing thin
37 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
38 that this functionality is critical to replicating cloud fractions comparable to those calculated from observations using Cloudnet.
39 UM_RA2T also has the extra source of sub-grid turbulent production of mixed-phase cloud; in a test with this option switched
40 off it was found that this process does not account for the improved cloud fraction comparison with observations (**Fig. S2**).
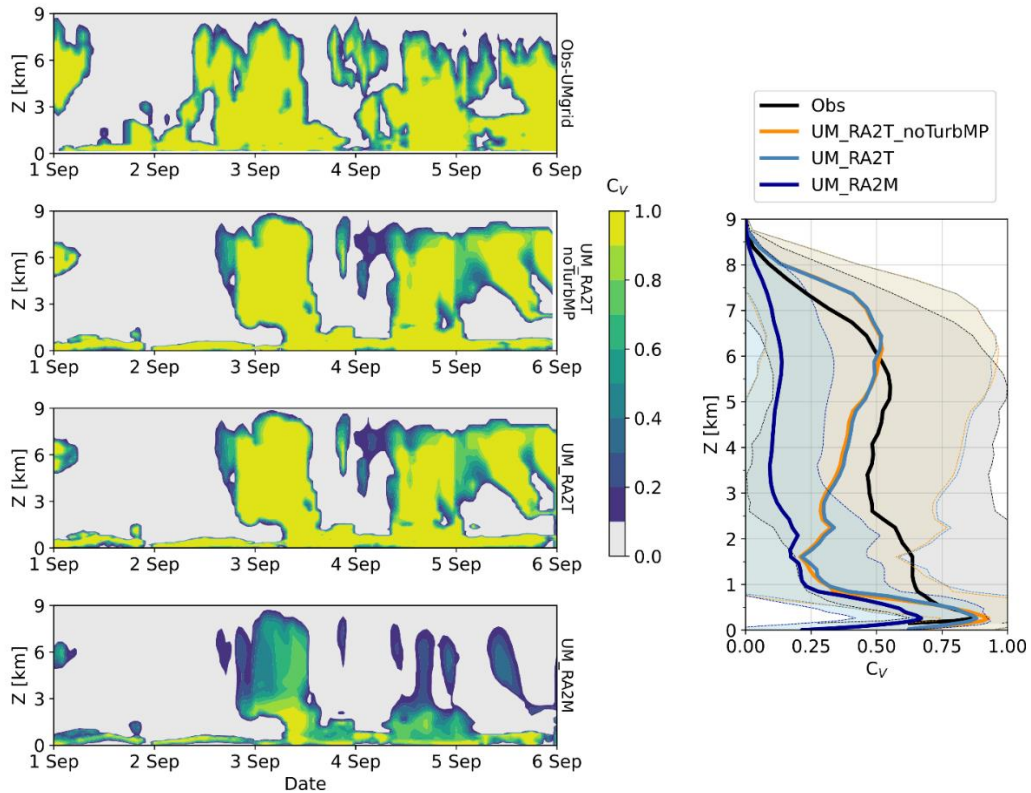
Figure S2: C_V (**a**) observed and modelled by (**b**) UM_RA2T without sub-grid turbulent production of mixed-phase cloud switched on (UM_RA2T_noTurbMP), (**c**) UM_RA2T, and (**d**) UM_RA2M for a subset of the drift period (1 Sep to 6 Sep), illustrating that the inclusion of this sub-grid process does not account for the high cloud fractions modelled by the UM_RA2T case.

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

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

54 In summary, our data highlight the difficulty of diagnostic cloud fractions, such as that computed by the **Smith (1990)** scheme,

55 to produce fully cloudy grid boxes on fine grid scales and indicate that models represent cloud fractions in different ways;

56 therefore, cloud water content analyses should be conducted in addition to cloud fraction comparisons when studying Arctic

57 clouds to provide a more robust model-observation comparison than using cloud fractions alone. **Figure S3** demonstrates how

58 using a TWC threshold to define in- and out-of-cloud regions can give a more consistent comparison between the observations

59 and models.

60 ## S1.2   Averaging observations to model grid

61 Liquid and ice water contents are calculated at Stage 2a of Cloudnet (**Illingworth et al., 2007**) using the various measured

62 inputs described in **Sect. 2.2** of this study and referencing the radar height grid and time resolution. For comparison with

63 numerical models, Cloudnet includes several additional functions in Stage 2b to average these observational data on to the

64 corresponding model grids in a consistent manner.

65 Observational data are split into each model grid box then statistics are calculated, e.g., grid-box mean liquid water content. To

66 ensure that there are enough data present in each box, a quality factor is applied: this factor is defined by default to be related

67 to 90% of the grid box size, designed at Cloudnet's creation to ensure that there were enough observational data within each >

68 10 km model grid box for meaningful statistical comparisons. However, we found that this high value for the quality-control

69 factor was too efficient in filtering out data with our higher spatial resolution (1.5 km) grid boxes.

70 This factor was reduced to 10% or 30% respectively for the IFS and UM grids utilised in this study to reduce the number of

71 profiles required for meaningful statistics within each 4D box. Code failures relating to too few data restricted the UM quality

72 factor being reduced further than 30%, but this is unsurprising given each box is 1.5×1.5 km in this study.
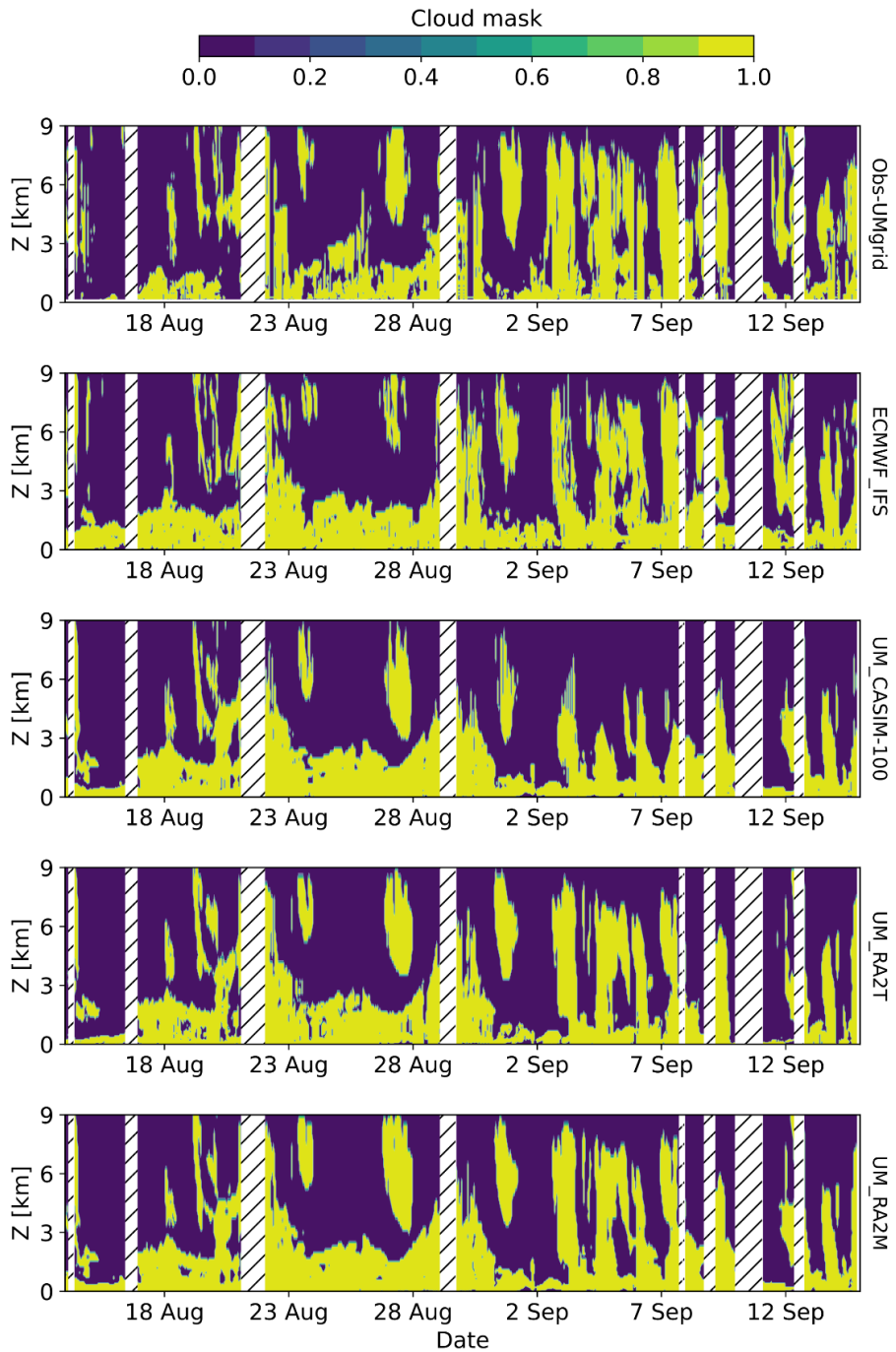
Figure S3: Comparison of cloud masks built using the in-cloud TWC threshold described in **Sect. 2.4**. TWC is calculated using all LWC and IWC data, then thresholding is applied.

73

74

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.
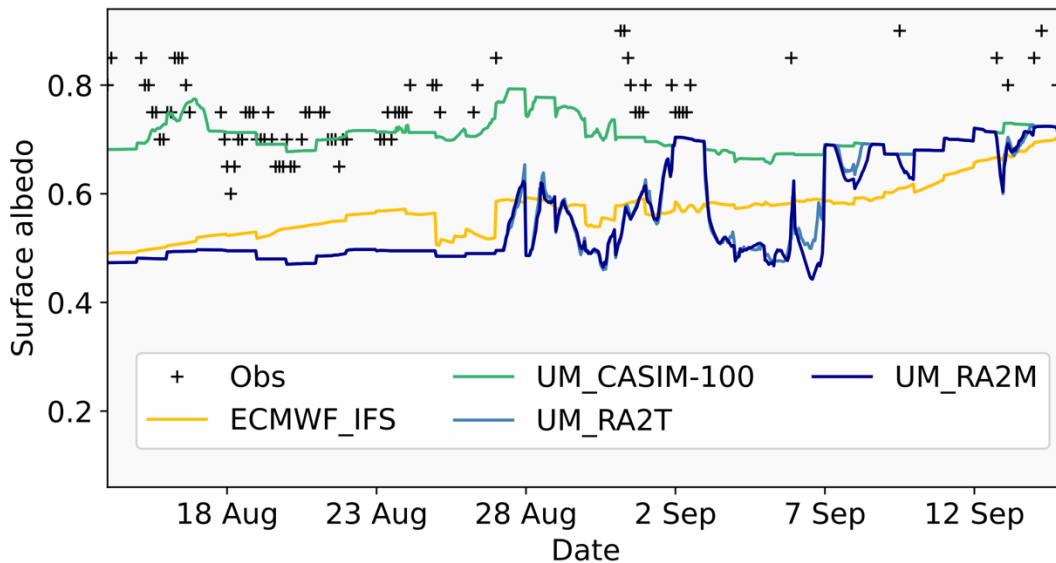

Figure S4: Surface albedo estimated from surface images of ice cover taken from the ship (**Obs**, black crosses) and diagnosed by the models.

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

94   begins to fall; however, our few observational data points during this period suggest that the models are still underestimating
95   albedo by approximately 10-15%.

96   While this comparison suggests that the models are performing poorly with regards to surface reflectivity, one must note that
97   the models are representing the albedo of a 1.5/9 km grid box (UM/IFS, respectively) while the observed estimates are taken
98   from the area immediately surrounding the ship. Therefore, any reduction in model albedo due to e.g., melt ponds or leads
99   would not be accounted for in our observational estimates.
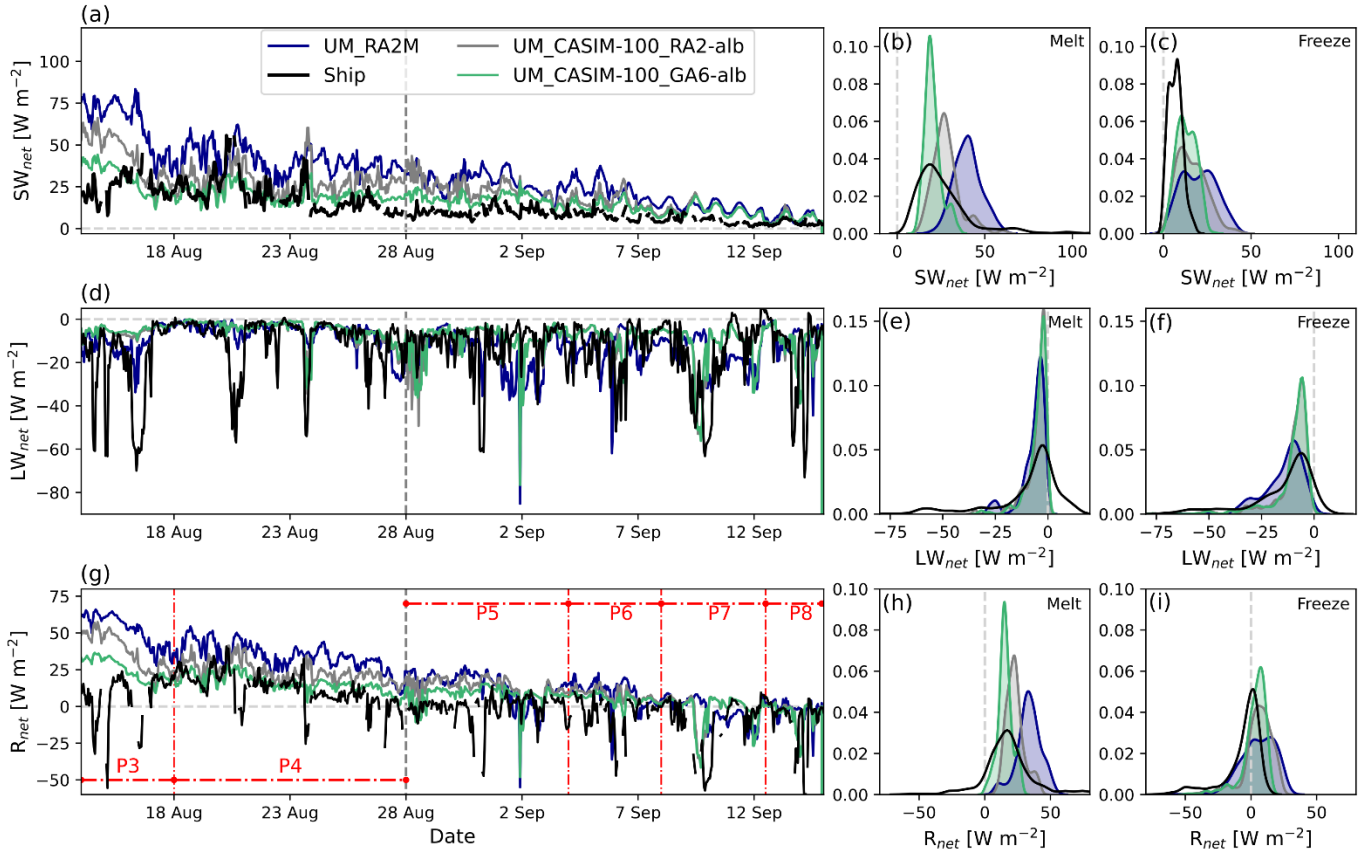
Figure S5: SW_{net}, LW_{net}, and R_{net} simulated by UM_CASIM-100 with albedo options for the Regional Atmosphere version 2 (UM_CASIM-100_RA2-alb; grey), updated Global Atmosphere version 6.0 (UM_CASIM-100_GA6-alb; green) used in the main body of this study, and UM_RA2M (dark blue) for reference. 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).

100   **Figure S5** shows the surface radiative balance modelled in UM_CASIM-100 (as **Fig. 2**, here labelled UM_CASIM-100_GA6-
101   alb) and UM_CASIM-100 with the default Regional Atmosphere limits for the surface albedo parameterisations used (labelled
102   UM_CASIM-100_RA2-alb, as used in UM_RA2M and UM_RA2T). **Figure S5** therefore shows that the cloud physics

103 representation of UM_CASIM-100 does still improve radiative interactions, with comparison with our observations, over
104 UM_RA2M (and UM_RA2T, not shown). Thus, the surface albedo updates are not the sole cause of its improved performance
105 over the operational UM schemes; however, the combination of the updated surface albedo (to represent snow on sea ice) and
106 improved cloud microphysical representation (from the CASIM scheme) yields the best UM comparison with observations (as
107 presented in the main body of this study). Also note that changing the surface albedo settings within the UM framework has
108 little effect on the modelled cloud microphysics (Fig. S6).
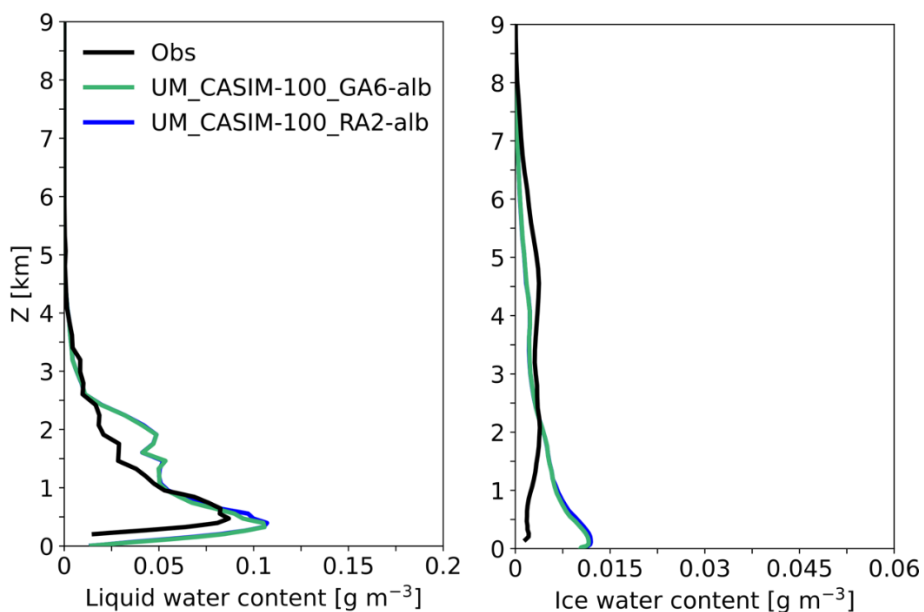
Figure S6: Mean liquid (*left*) and ice (*right*) water contents modelled by the UM CASIM simulations with the updated GA6 albedo settings (green, as UM_CASIM-100 in the main body of the study) and the default RA2 albedo settings (as described in this section). Observations are again shown in black, as Figure 4.

109 **S3   Point versus swath comparison**

110 This study centres on the extraction of the ship's location hourly from the model datasets considered; it is a direct, one-to-one
111 comparison between the model fields and the measurements. However, such model evaluations can benefit from considering a
112 swath of data over a given area surrounding the observations to account for spatial and temporal heterogeneity in the model
113 fields.

114 In this section, we show how such an approach may influence the main findings of this study, using the UM_CASIM-100 data
115 as illustration. Two swath areas were designed dependent on the location of the ship during its drift to coincide with the
116 changing ship track (shown in **Fig. S7**): the first (in black) is used between 14 Aug (start of drift) and 2 Sep, while the second
117 (in grey) is used between 3 Sep and 14 Sep (end of drift). As such, the first swath covers 21×50 grid boxes and the second
118 covers 72×20 grid boxes in the LAM grid.

Figure S7: Two swath areas designated for model-observation comparisons. (1) Black hatched region: 14 Aug – 2 Sep, (2) Grey hatched region: 3 Sep – 14 Sep. Ship track colour changes from orange to red at 3 Sep to aid interpretation.

119   C_V, q_{liq}, and q_{ice} data extracted from UM_CASIM-100 using the point and swath comparison methods are shown in **Figure**
120   **S8**. In contrast to the main body of this study, we show median profiles over the drift (supplemented by ± one standard deviation
121   in shading) for fair comparison as the swath analysis considers far more data points than the point comparison (thereby
122   potentially influencing the mean more so than the median of the distribution).

123   C_V and q_{liq} vary only subtly between the two data extraction methods and between the two seasons; however, somewhat larger
124   differences between the point and swath comparisons are evident in the q_{ice} data during the freeze period (**Fig. S8**). Between
125   1.5 and 9 km, the swath extraction method produces a median q_{ice} that is consistently smaller than that which was achieved
126   using the point comparison, though it is important to note that the median modelled q_{ice} is low in both data extraction scenarios.
127   Modelled cloud ice has a greater degree of spatial and temporal heterogeneity than the cloud fraction or cloud liquid and is
128   therefore affected by the amount of data included in the sample for consideration. During the melt period, the cloud ice content
129   is consistently low, but non-zero, and this "background" state of low cloud ice content continues during the freeze period.
130   However, with colder air temperatures during the freeze, there are also more sporadic occurrences of higher q_{ice}, acting to
131   increase the standard deviation of the data at most heights. The increased sample size in the swath analysis therefore acts to
132   slightly reduce the median cloud ice content and its variability (over one standard deviation) with comparison to the point
133   comparison, as the sporadic high ice content clouds are less common than the "background" low ice content clouds.

134   This additional perspective highlights that the cloud fraction and cloud liquid water content analysis included in the main body
135   of this study are robust; however, the heterogeneity of the ice phase in these clouds affects the q_{ice} comparisons and must be
136   approached with some caution. As the ice phase was not the focus of our study, and we have noted its variability in our analysis,
137   the conclusions reached from the point comparisons of this study would be unaffected by the use of the swath, rather than the
138   point, comparison.
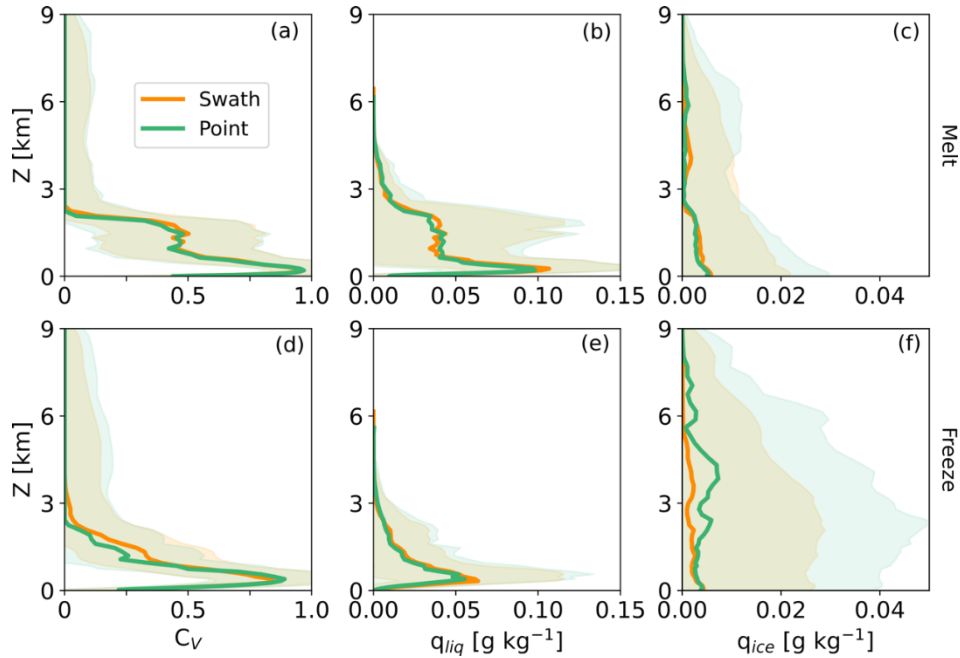
Figure S8: Comparison of C_V, q_{liq}, and q_{ice} data from UM_CASIM-100 over the melt (*a—c*) and freeze (*d—f*) period of the ship's drift. Median profiles of each quantity are shown, with ± one standard deviation in shading. (*a, d*): C_V, (*b, e*): q_{liq}, and (*c, f*): q_{ice}.

S4   Error propagation from GLM to LAM

In Section 3.3.1, we discuss how the initial conditions of the driving model can affect the thermodynamic biases of the UM LAM. A test simulation (UM_RA2M-ERAI-GLM) was conducted using ERA-Interim to initialise the global model instead of the Met Office analyses over a subset of the drift period, and **Figure 11** shows that the lower atmosphere behaves more like ECMWF_IFS than the UM_RA2M simulation.

Figure S9 shows the temperature and moisture biases from the two global model initial condition datasets considered: the UM global analyses and ERA-Interim. Note that only 6 profiles are shown (corresponding to 1200 UTC) for each dataset in **Figure S9**, in contrast to **Figure 11** which includes 6-hourly data (24 profiles in total). A key point to note is that both initial condition datasets lack strong thermodynamic biases below 2 km; biases such as that exhibited by both UM_RA2M and UM_RA2M-ERAI-GLM. This lack of strong low altitude thermodynamic biases in the initial conditions lends weight to our conclusion that the model physics (here, specifically, clouds) is driving the thermodynamic biases below approximately 2 km. However, the UM analyses do exhibit increasing temperature biases with height and ERA-Interim exhibits a cold bias between 2 km and 8 km, both of which are mirrored somewhat in the UM_GLM (and UM_RA2M and UM_RA2M-ERAI-GLM) and ECMWF_IFS biases presented in **Figure 11**.
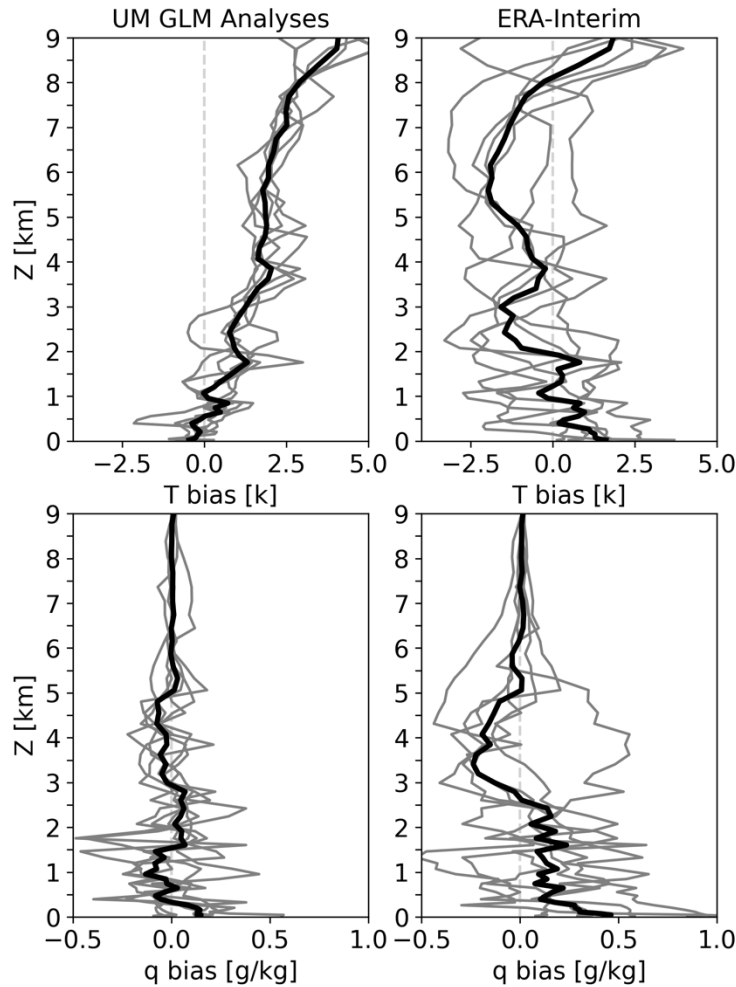
Figure S9: Biases of the UM global analyses (*left*) and ERA-Interim (*right*) with respect to radiosondes launched at 1200 UTC between 31 Aug and 5 Sep (aligning with the period discussed in Figure 11). Top row: temperature biases; bottom row: moisture biases. The median of the 6-day period is shown in black, with individual daily bias profiles (at 1200 UTC) of each initial condition dataset shown in grey.

154 As lateral boundary conditions for the LAM are generated hourly from the global model, there is not sufficient free-running
155 time to assess how the LAM biases would evolve without the influence of the driving model, and thereby evaluate how much
156 of the thermodynamic biases can be attributed to the global model. Figure S10 shows the temperature biases exhibited by three
157 example 36-h forecasts (20180815T1200Z, 20180828T1200Z, and 20180902T1200Z) to illustrate, on a forecast-by-forecast
158 basis, how the LAM simulations differ from the global model in their thermodynamic structure. These forecasts were chosen
159 as they provide a range of biases with respect to the radiosondes – 20180815 has strong temperature biases, 20180828 has only
160 weak biases, and 20180902 has a mixture of both.
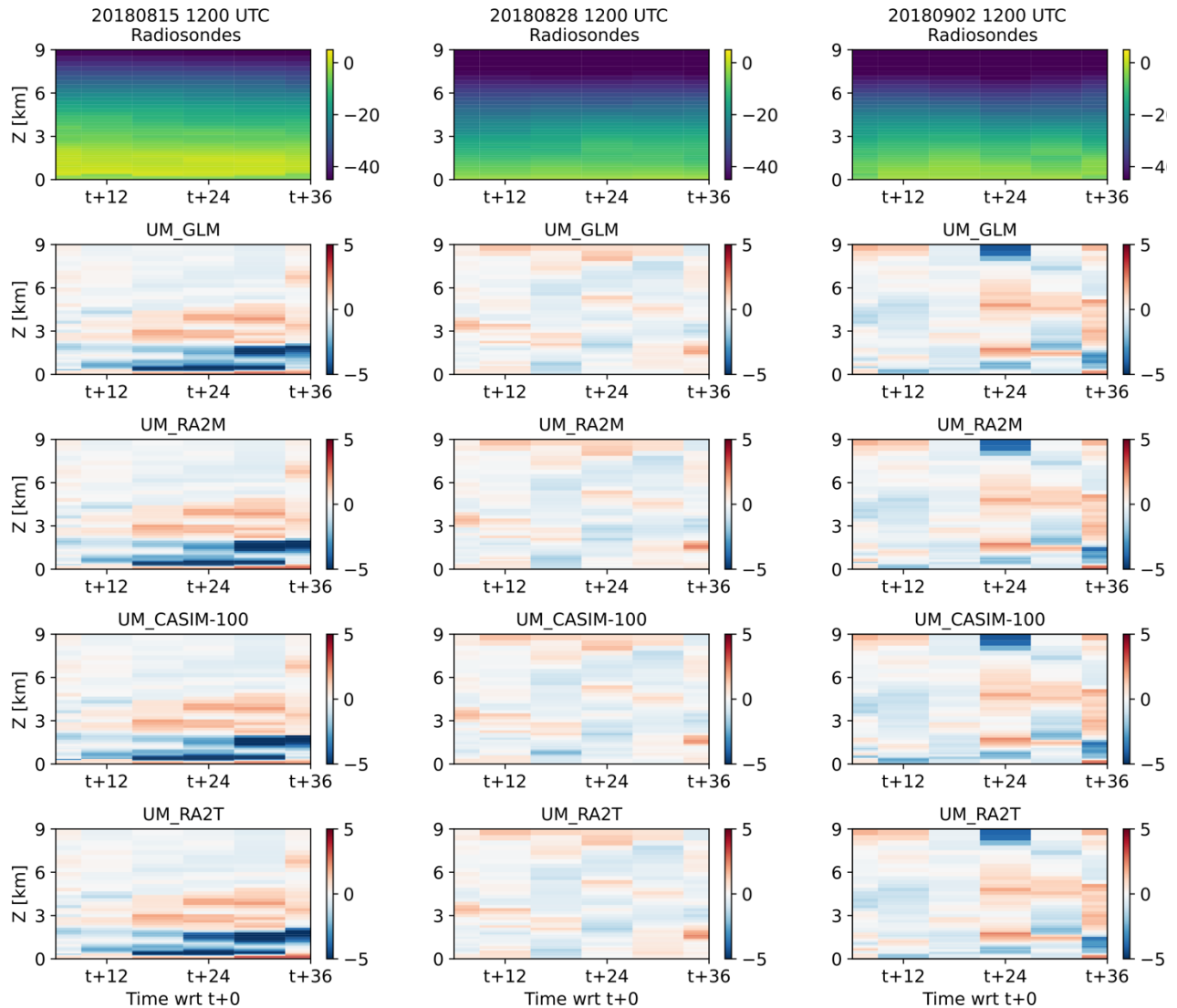
Figure S10: Three examples of 36-h forecasts, with radiosondes during the given time period shown on the top row for reference. Data from hours 6-36 are shown. Biases with respect to these radiosonde data are shown in the subsequent panels. *Second row*: UM_GLM; *third row*: UM_RA2M; *fourth row*: UM_CASIM-100, and *bottom row*: UM_RA2T. *Left*: 20180815T1200Z forecast; *middle*: 20180828T1200Z forecast; and *right*: 20180902T1200Z forecast.

161

162 **Figure S10** illustrates that the three LAM simulations (bottom three rows) have a strong resemblance to the UM_GLM biases
163 (second row) over each of these example forecasts. However, there is more structure in the differences between the UM_GLM
164 and the LAMs than **Fig. S10** suggests. As an example, **Fig. S11** shows individual bias profiles of temperature, moisture, and
165 cloud liquid water mixing ratio from each LAM simulation with respect to the global model for the 20180815T1200Z forecast.
166 At each time interval shown, biases exist in all three parameters; the biases from the LAM simulations do differ from the global
167 model. In the 20180815T1200Z forecast, biases with respect to the global model (and radiosondes) do increase with time

168 through the forecast; however, this is not true for every forecast (for example, the 20180828T1200Z or 20180902T1200Z

169 forecasts shown here; **Fig. S10**).

170 From **Fig. S11**, it can also be noted that there are clear occasions where the LAM boundary layer depth differs from the global

171 model, as evidenced by the strong positive and negative peaks in temperature, moisture, and/or q_{liq} biases at adjacent altitudes.

172 However, a simple misplacement of the boundary layer inversion does not explain the strong temperature and moisture biases

173 between 1 and 2.5 km at t+24 and t+30 h – these higher altitude biases are strongest in the UM_RA2T simulation, indicating

174 that they may relate to e.g., the higher cloud ice content reproduced by this simulation with comparison to UM_CASIM-100

175 and UM_RA2M.
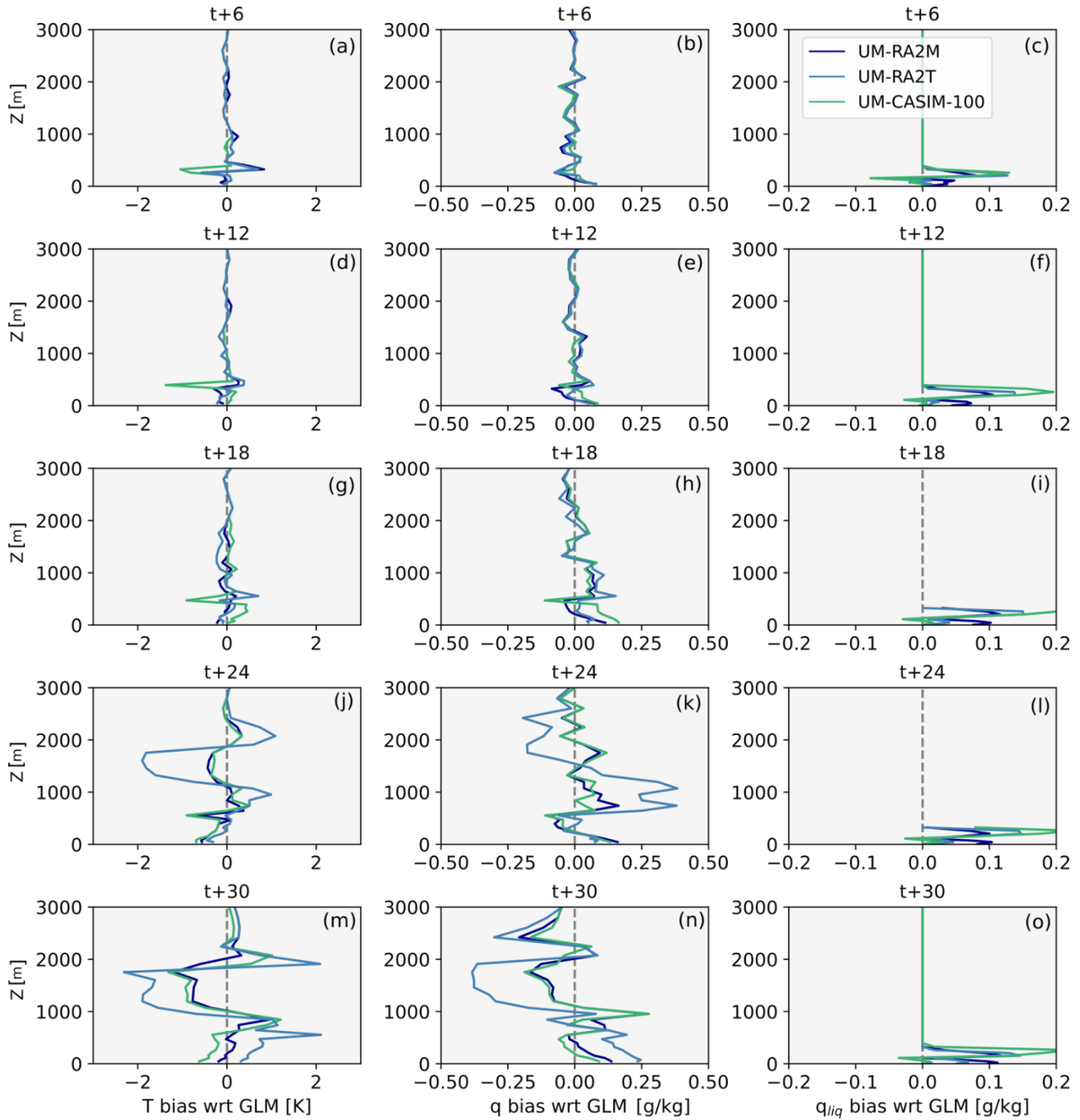
Figure S11: Temperature (*left*), moisture (*middle*), and cloud liquid water mixing ratio (*right*) biases of each LAM simulation with respect to the UM global model for hours 6—30 of the 20180815T1200Z forecast. (a—c): t+6, (d—f): t+12, (g—i): t+18, (j—l): t+24, and (m—o): t+30.

177 **S5   UK Chemistry and Aerosol (UKCA) model**

178 UKCA simulates gas and aerosol chemistry and transport in the atmosphere using the GLObal Model of Aerosol Processes

179 (GLOMAP-Mode, **Mann et al., 2010**) and an atmospheric chemistry scheme, with an additional boundary layer nucleation

180 scheme used to simulate gas-to-particle conversion of sulphuric acid to sulphate aerosol (**Spracklen et al., 2010**). To generate

181 the aerosol input files for CASIM, UKCA was one-way coupled to the UM at version 11.2 using the Global Atmosphere 7.1

182 dynamical core (**Walters et al., 2019**). Daily averaged soluble accumulation- and coarse-mode aerosol number and mass

183 concentrations calculated from UKCA grid points north of 88.125 °N (**Fig. S2**) were used as input to the UM.
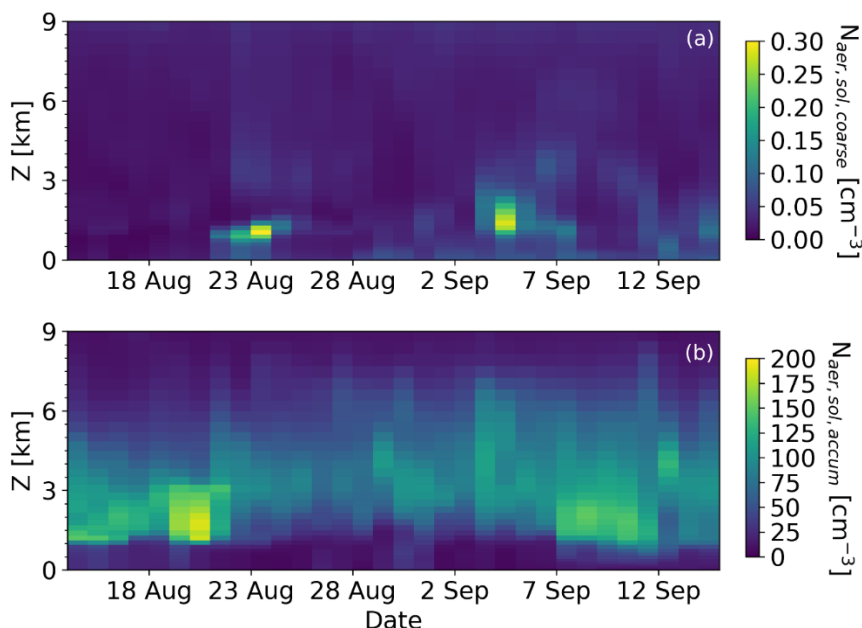

Figure S12: Soluble (**a**) coarse- and (**b**) accumulation-mode aerosol number concentrations simulated by UKCA and used as input
for the CASIM scheme in the UM_CASIM-AeroProf simulation. Aerosol profiles are kept constant over each daily forecast, with no
aerosol processing by cloud.

184 **S6   Periods of consistent meteorology**

185 To better understand how the model thermodynamic biases relate to cloud properties in each simulation, we split our drift

186 period further into four subsections – periods 3 to 6, as illustrated in **Figs. 2 and 8** – to study periods of consistent meteorology.

187 Mean equivalent potential temperature (θ_e) and q profiles measured by radiosondes during these periods are shown in **Fig. 12**.

188 Of the four periods considered, period 3 had cloud-free conditions most often. Periods 5 and 6 were similar; both were cloudy

189 and influenced synoptically by three different low-pressure systems over their duration.

190
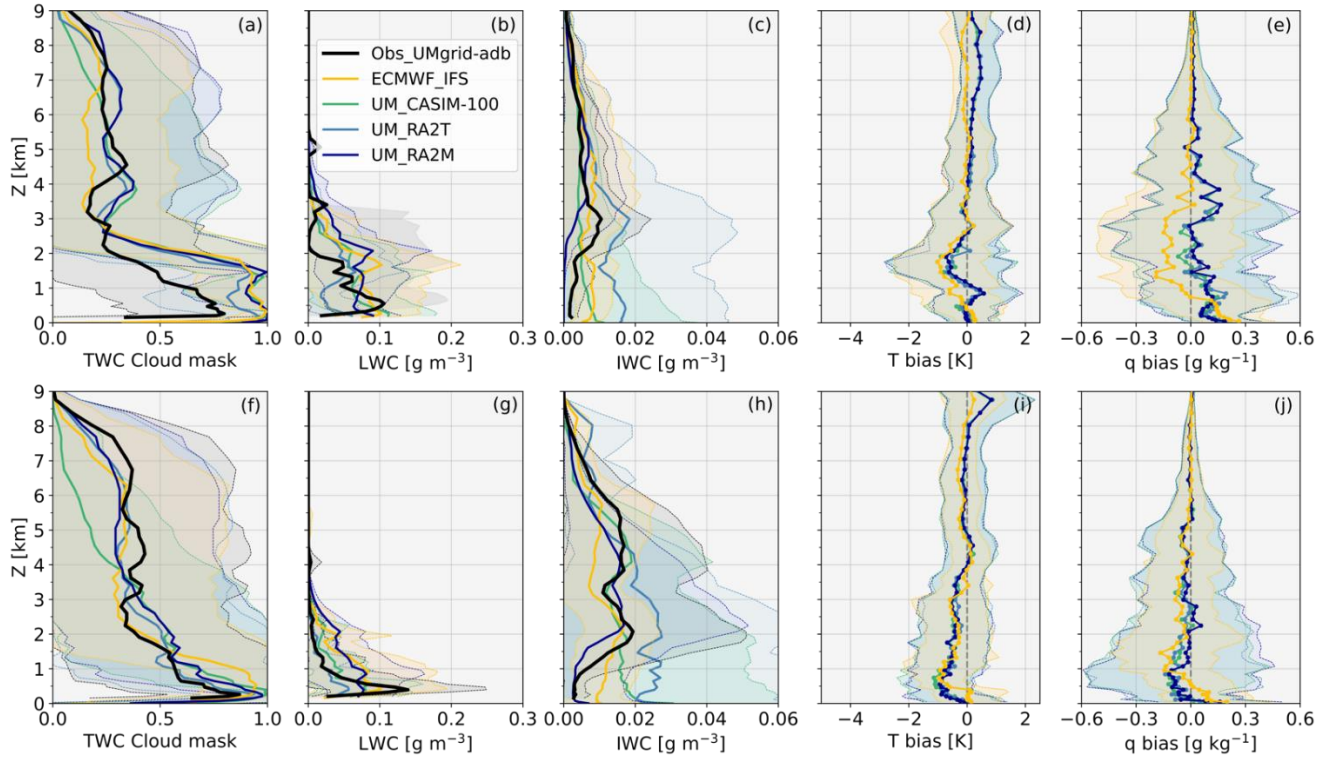
Figure S13: (As **Fig. 13**) Comparison of mean cloud mask, LWC, and IWC profiles with median biases in T and q with respect to radiosondes for period 4 (**a—e**, *top row*) and period 5 (**f—j**, *bottom row*). Again, observed LWC calculated assuming adiabatic conditions using Cloudnet. ± one standard deviation shown in shading to illustrate variability.

191   **Figure S13** shows comparisons of the TWC cloud mask, in-cloud LWC and IWC, and associated T and q biases over periods

192   4 and 5. Both periods 4 and 5 support the findings of periods 3 and 6. During period 4, the models overpredict cloud occurrence

193   below 2 km similar to period 3; however, both the LWC and IWC are in better agreement with observations during period 4.

194   Similarly, all simulations agree better with observed cloud occurrence during period 5 (consistent with our result for period 6),

195   and both the LWC and IWC are again in reasonable agreement during this time window. Consequently, the model

196   thermodynamic biases with respect to radiosonde measurements are weaker (though still present) during periods 4 and 5 than

197   during periods 3 and 6.

198   **S7**   **Primary ice nucleation parameterisation**

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

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

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

202   smallest, at e.g., –10 °C. To test whether the method of parameterising primary ice itself has any effect on these biases, we

203   trialled the use of each of the **Fletcher (1962)**, **Cooper (1986)**, and **Meyers et al. (1992)** parameterisations within the CASIM

204   framework; however, we found little difference in our tropospheric ice due to the different parameterisation methods.

205   Changing the primary ice parameterisation alters biases slightly within the lowest 3 km of the domain, with a maximum

206   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

207   (**Figure S14**). Differences shown here are much smaller than the more significant UM configurations changes/IFS comparison

208   shown in **Figs. 9, 11, 13**.
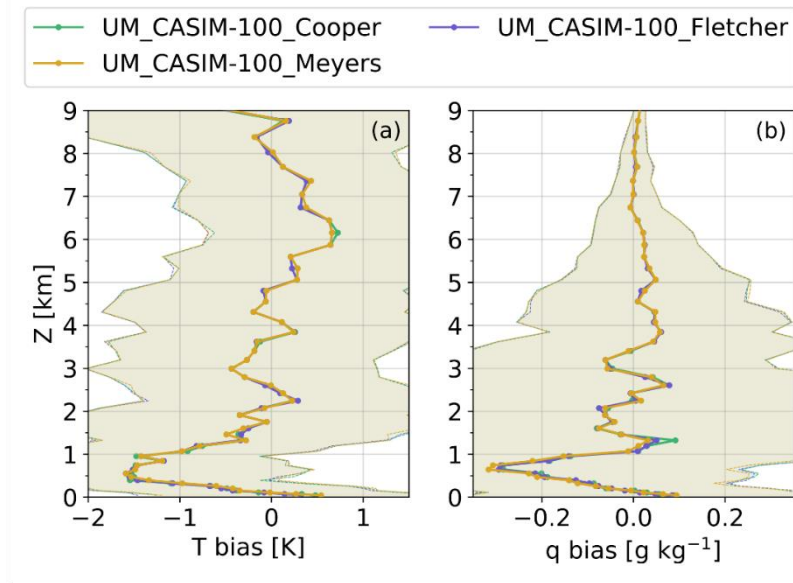
Figure S14: Comparison of T and q biases (with respect to radiosonde measurements) of CASIM-100 runs with different primary ice production parameterisations imposed (green: **Cooper, 1986**; purple: **Fletcher, 1962**; gold: **Meyers et al., 1992**), over the drift subset of 31 Aug to 5 Sep. Here, UM_CASIM-100_Cooper is equivalent to UM_CASIM-100 data shown in the main body of the paper.

209

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.

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

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

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

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