Article:
Van Weverberg, K., Giangrande, S., Zhang, D. et al. (2 more authors) (2023) On the Role of Macrophysics and Microphysics in Km-Scale Simulations of Mixed-Phase Clouds During Cold Air Outbreaks. Journal of Geophysical Research: Atmospheres, 128 (11). e2022JD037854. ISSN: 2169-897X

https://doi.org/10.1029/2022jd037854

On the Role of Macrophysics and Microphysics in Km-Scale Simulations of Mixed-Phase Clouds during Cold Air Outbreaks

K. Van Weverberg[1,4,5], S. Giangrande[2], D. Zhang[3], C. J. Morcrette[1,6], P. R. Field[1,7]

[1]Atmospheric Processes and Parametrizations, Met Office, Exeter, United Kingdom
[2]Brookhaven National Laboratory, Upton, New York, United States
[3]Pacific Northwest National Laboratory, Richland, Washington, United States
[4]Department of Geography, Ghent University, Ghent, Belgium
[5]Royal Meteorological Institute of Belgium, Brussels, Belgium
[6]Global Systems Institute and Department of Mathematics, University of Exeter, Exeter, United Kingdom
[7]School of Earth and Environment, University of Leeds, Leeds, United Kingdom

Key Points:

- The role of cloud-fraction parameterizations in simulations of mixed-phase clouds during cold-air outbreaks is investigated.
- Cloud-fraction parameterizations matter for cloud-radiative effects, but have limited impact on the liquid water content.
- Simulations with reduced snow depositional growth improve liquid water path and precipitation size distributions.

Corresponding author: Kwinten Van Weverberg, kwinten.vanweverberg@ugent.be

Abstract

Regional atmospheric models struggle to maintain supercooled liquid in mixed-phase clouds during polar cold-air outbreaks (CAOs). Previous studies focused on the parameterization of aerosol, microphysics and turbulence to understand the origin of this widespread model bias. This study investigates the role of macrophysics parameterizations (MacP) in the simulation of mixed-phase clouds. Km-scale simulations are performed for a large number of CAO cases over Norway, for which continuous ground observations were collected at one site over 6 months. We use a novel analysis that attributes the cloud-radiative errors to deficiencies in specific cloud regimes. We show that the MacP matters for cloud-radiative effects in CAOs, but that it is probably not the primary cause of the lack of liquid water in simulated mixed-phase clouds. Of all the MacP sensitivities explored in this study, the prognostic representation of both liquid and ice shows most promise in increasing the liquid water path. A newly proposed hybrid MacP with prognostic frozen and diagnostic liquid cloud fraction reproduces some of the benefits of the prognostic scheme at reduced cost and complexity. The two-moment microphysics scheme in this study produces too large precipitation particles. Reducing the snow deposition rate decreases the precipitation particle sizes and largely improves the liquid water path. Simulations are less sensitive to reduced riming rates.

Plain Language Summary

Cloud droplets and ice crystals often coexist in polar clouds, even at temperatures well below the freezing level. These clouds play an important role in the climate system, since they reflect more sunlight back to space than completely frozen clouds. Therefore it is important that climate and weather forecasting models are thoroughly evaluated in their ability to produce these so-called mixed-phase clouds. This study compares simulations of polar clouds with detailed cloud observations over Norway. We confirm that simulations produce clouds that are nearly completely frozen, while observations clearly show both liquid and frozen clouds. We test sensitivities of the simulations to specific model components to assess potential avenues for model improvement. Of all the tests performed, the ones affecting the simulated snowflake sizes seem most promising in allowing both liquid and frozen cloud to coexist.

1 Introduction

Mixed-phase clouds in polar cold-air outbreaks (CAOs) continue to pose a major challenge for regional atmospheric models (Klein et al., 2009; Abel et al., 2017; Field et al., 2017). The co-existence of ice and supercooled liquid in these clouds at temperatures well below the freezing point, challenges our understanding of the interaction between aerosol, microphysical and turbulent processes. Yet, it is imperative that progress be made in the representation of mixed-phase clouds, given their abundance over polar regions and their potentially important cloud-radiative feedbacks in a changing climate (Frey & Kay, 2018; Murray et al., 2021). Despite progress in the representation of the interaction of turbulent processes and microphysics (Korolev & Field, 2008; Field, Hill, et al., 2014), most atmospheric models still struggle to reproduce the observed amount of supercooled liquid in polar mixed-phase clouds (Klein et al., 2009; Abel et al., 2017; Field et al., 2017).

Scales involved in the maintenance of this supercooled liquid range from the microphysical scale, to sub-cloud scale turbulent eddies, and open- and closed-cell mesoscale organization. This implies that a number of parameterizations are involved in the correct simulation of mixed-phase clouds. Solomon et al. (2009) and Klein et al. (2009) showed that two-moment microphysics parameterizations capture the supercooled liquid water content in the mixed-phase clouds better than one-moment schemes. Too rapid glacia-

tion through riming processes was found to deplete liquid water effectively in mixed-phase clouds over the Southern Ocean simulated by Furtado and Field (2017). Abel et al. (2017) pointed to the importance of precipitation formation and autoconversion from cloud to rain as an important driver for liquid water depletion. In a large intercomparison of convection-permitting CAO simulations, Field et al. (2017) found all simulations to struggle to reproduce sufficient supercooled liquid. It was found that the dynamical core, and the convective and boundary-layer parameterization, played an important role in the evolution of the cloud fields. In another study, Field, Cotton, et al. (2014) demonstrated that the significant underestimation of supercooled liquid in a CAO case over northern Europe could be alleviated by changing the boundary-layer diagnosis and inhibiting the ice formation at warm temperatures.

It has often been suggested that phase separation plays an important role in the maintenance of supercooled liquid in CAO clouds, even at km-scale. This can be related to continuous turbulent generation of liquid in updraft regions, while depositional growth of the ice phase is limited to downdraft regions (M. Shupe et al., 2008; Korolev & Field, 2008). Another often cited mechanism is phase separation through sedimentation of ice particles near the cloud top, while condensation of liquid takes place through longwave cooling (Rauber & Tokay, 1991; Forbes & Ahlgrimm, 2014). This phase separation typically occurs at subgrid scales and hence surprisingly few studies have been devoted to the role of macrophysics parameterization (MacP) in convection-permitting simulations of CAO clouds. One exception is the study by Abel et al. (2017) showing that a reduced subgrid overlap between the liquid and ice phase in a MacP improved their simulation of a CAO event over the United Kingdom. While many km-scale models ignore subgrid cloud variability, several studies find that the inclusion of a subgrid MacP in these models is beneficial for cloud-radiative processes (Hughes et al., 2015; Boutle et al., 2016; Bush et al., 2020; Van Weverberg, Morcrette, & Boutle, 2021).

Macro- and microphysics are closely intertwined. Indeed, the macrophysical treatment of ice and snow growth in microphysics parameterizations (MicP), i.e. partitioning of grid boxes in ice- and liquid sub- and supersaturated regions, also affects the coexistence of liquid and ice in polar clouds.

In this paper, we explore the role of MacPs, and of ice growth in MicPs, in mixed-phase cloud evolution using the Regional Atmosphere and Land 3 (RAL3) configuration of the UK Met Office Unified Model (UM). Km-scale simulations are performed with a range of MacP configurations for 47 CAO cases identified during the Cold-Air Outbreaks in the Marine Boundary Layer Experiment (COMBLE, Geerts et al. (2022)). The continuous ground-based observations collected at one site over a 6 month period provide a unique opportunity to perform a more statistically robust analysis of the cloud phase compared to single case studies observed by aircraft. Sensitivities explored in this paper include (1) the formulation of the subgrid saturation-departure distribution in diagnostic MacPs, (2) the role of the subgrid phase-overlap between the liquid and the ice phase, and (3) the added value of using a prognostic rather than diagnostic representation of liquid and frozen fractions. The latter experiments include a novel hybrid MacP approach with prognostic ice, but diagnostic liquid cloud fractions. Furthermore, (4) the role of snow deposition and riming in the microphysics parameterization (MicP) is investigated. Particular attention is given to the collocated radiation and cloud property biases, identifying those cloud regimes contributing most significantly to cloud-radiative errors.

The remainder of the paper is organised as follows. Section 2 provides some background on the COMBLE field campaign and lists the observations used in this study. The sensitivity experiments are explained in Section 3. Section 4 discusses all the analysis performed with these simulations and a summary is provided in Section 5.

2 COMBLE field campaign and Observations

2.1 COMBLE field campaign

From 1 December 2019 to 1 June 2020, the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Mobile Facility (AMF; Miller et al. (2016)) was deployed at a coastal site in northern Norway (near Andenes, 69.14N, 15.68E, further referred to as ANX), in the framework of COMBLE (Geerts et al., 2022). During this six-month campaign, there were a total of 47 CAO events that lasted from hours to days, leading to a cumulative total of 34 days of CAO conditions. CAO events were identified following Geerts et al. (2022), based on the following observed conditions: A potential temperature contrast between the sea surface temperature (SST) and the 850 hPa pressure level ($\Theta_{SST}-\Theta_{850}$; further referred to as $\Delta_{SST-\Theta 850}$) >0, a surface wind speed >10 kt, and a wind direction between 250° and 30°. A list of all CAO cases during the COMBLE campaign can be found in the Supplementary Material of Geerts et al. (2022). CAO events can broadly be categorized as one of two distinct modes. A first mode consists of more convective events with heavier precipitation and alternating strong updrafts and decaying convective cells (Mode A as described in Geerts et al. (2022), further referred to as the convective (CNV) mode in this study). The second mode is more stratiform and quiescent with weaker updrafts, lighter rain and lower cloud tops (Mode B as described in Geerts et al. (2022), further referred to as the stratiform (STR) mode in this study). The first, unstable mode is typically associated with larger $\Delta_{SST-\Theta 850}$ than the second mode. In this paper, we analyze both modes separately where CNV and STR cases have $\Delta_{SST-\Theta 850}$ >2 K and <2 K, respectively.

2.2 Observations

2.2.1 MicrobaseKaPlus: Liquid water content, ice water content and cloud fractions

Vertically distributed cloud locations and water contents are derived from the MicrobaseKaPlus ARM synergistic retrieval (further referred to as MBase). MBase provides profiles of liquid and ice water content, based on ceilometer, micropulse lidar, microwave radiometer, balloon soundings and vertically pointing Ka-band cloud radar (KAZR) data at the ANX site with high temporal (4s) and vertical (30m) resolution (M. Wang et al., 2022). To compare observed vertical profiles for one location with a simulated grid column, the observations are averaged over the time it takes an air parcel to cross one model grid length of 1 km. The observed wind speeds at each model level are used to obtain 'grid-box' averaged liquid and ice water content (q_{liq} and q_{fro}) and an 'observed' cloud fraction (CF). This approach has been widely used in model evaluation studies (Illingworth et al., 2007; Morcrette et al., 2012; Van Weverberg et al., 2015) and gives reliable results for low-order moments and long time series (Gruetzun et al., 2013). Zhao et al. (2014) estimated the random uncertainties in the retrievals of q_{liq} and q_{fro} to be around 15 and 55% respectively. MBase is also used to obtain vertically integrated liquid and ice water paths (LWP and IWP). Note that the LWP from MBase is constrained by the microwave radiometer retrievals with an uncertainty of about 30 gm^{-2} (Turner et al., 2007). The presence of supercooled liquid water and drizzle poses additional challenges to the MBase retrievals (Cadeddu et al., 2020). Therefore, these retrievals are used as a baseline, but we use a more advanced classification to evaluate the vertical phase distribution, as explained in the next section.

2.2.2 ThermoCldPhase: Cloud phase classification

Apart from MBase, an independent algorithm is used for the identification of the cloud thermodynamic phase. The ARM ThermoCldPhase retrieval (Zhang & Levin, 2022), further referred to as TPhase, provides vertically-resolved cloud and precipitation phase

classifications using coincident lidar, radar, and radiometer measurements following the multi-sensor method developed by M. Shupe (2007). Hydrometeors are classified into liquid, drizzle, liquid + drizzle, rain, ice, snow, and mixed-phase based on lidar back-scatter intensity (or its gradient) and depolarization ratio, radar reflectivity, mean Doppler velocity, and Doppler spectrum width measurements. Ancillary measurements of liquid water path from the microwave radiometer and temperature profiles from radiosonde measurements are used to further constrain cloud hydrometeor classifications. Once cloud hydrometeor phases are determined, the whole cloud layer thermodynamic phase is classified as liquid, mixed-phased, or ice following the method by Korolev et al. (2017). If the fraction of ice-containing hydrometeors (e.g., ice, snow, and mixed-phase) in the whole cloud layers (μ_{ice}) is smaller than 0.1, the cloud layer is determined as a liquid layer. If $0.1 < \mu_{ice} < 0.9$, it is a mixed-phased cloud layer; and if $\mu_{ice} > 0.9$, it is an ice cloud layer. The time-averaging method described in section 2.2.1 was applied to TPhase using liquid, mixed and ice definitions above. The classification thresholds and the use of constraints from ancillary measurements are based on understanding of cloud hydrometeor physical properties (e.g., particle size distribution, shape, number concentration, falling velocity, freezing/evaporation temperatures) and from the literature (Eloranta, 2005; M. Shupe et al., 2004, 2006). The multi-sensor cloud thermodynamic phase classification method has been widely used for studying various cloud phase transition processes (Kalesse et al., 2016; Persson et al., 2017; Silber et al., 2018) and long-term cloud thermodynamic phase characteristics (M. D. Shupe, 2011) and for evaluating cloud and climate model simulations (Solomon et al., 2011; Zhang et al., 2019). Quantitative assessment of the uncertainty of the classification algorithm is difficult due to the lack of a definitive validation dataset. Although the ground-based multi-sensor method has not been validated, a similar method using multi-sensor airborne lidar and radar measurements to identify hydrometeor particle types was compared and evaluated against aircraft in situ cloud probe measurements, which shows good agreement in over 70% of cases (Romatschke & Vivekanandan, 2022). As guidance to the sensitivity of the time-averaging method, we provide observed cloud-property profiles averaged using a range of $0.5-2$ times the observed wind speeds in the following analysis. For COMBLE, the TPhase data are not available before February 11, 2020 because the Micropulse lidar (MPL) data are missing at the beginning of the campaign.

2.2.3 Vertical Profiles of wind speeds, temperature and humidity

To obtain observed cloud fractions from MBase and TPhase, the interpolated sounding retrieval was used (Jensen et al., 2022), which interpolates observed soundings to a regular time-height grid with 332 levels and a 1-minute resolution. No first guess model data enter this retrieval and the humidity profiles are adjusted between soundings using the microwave radiometer measurements. Typically, four soundings were launched daily during COMBLE. Uncertainties vary with atmospheric conditions. Sharp fronts might be missed by the interpolation, although most of the cases in this study are well within cold air masses behind cold fronts, with fairly homogeneous atmospheric conditions. Interpolated soundings are also used to calculate the observed $\Delta_{SST-\Theta 850}$ and the Estimated Inversion Strength (EIS; Wood and Bertherton (2006)):

$$EIS = LTS - {\Gamma_m}^{850}(Z_{700} - LCL) \tag{1}$$

Where LTS is the lower tropospheric stability ($LTS = \Theta_{700} - \Theta_{surf}$; Klein and Hartmann (1993)), ${\Gamma_m}^{850}$ is the moist-adiabatic potential temperature gradient at 850 hPa, Z_{700} is the height of the 700 hPa level and LCL is the lifting condensation level (calculated following Romps (2017)).

2.2.4 Surface Radiation

Surface radiation observations are obtained from the ARM Best Estimate Cloud and Radiation Data product (Xiao & Xie, 2022). Radiometer surface downwelling longwave (LW) and shortwave (SW) radiation was used, with a temporal resolution of 60 s. This product averages two out of three different co-located instruments measuring irradiances that agree best with each other (Shi & Long, 2002). Uncertainties are of the order of 6 and 2.5% for SW and LW radiation respectively (Stoffel, 2005).

2.2.5 Surface Precipitation

Reliable surface precipitation data in a site like Andenes are hard to obtain, given the prevalence of mixed-phase and frozen precipitation. Many of the radar retrievals for quantitative precipitation estimates become too uncertain to be used for model evaluation. This study uses both the weighing bucket (WB) and the Present Weather Detector (PWD) rain gauge data from the Surface Meteorological Instrumentation (Kyrouac & Shi, 2022), aggregated from the native resolution of 1 minute to 15 minute intervals. Uncertainties in these instruments are typically of the order 0.1 mm or 10% of the reported instantaneous value, whichever is larger (Jia et al., 2020).

3 Simulations

3.1 General Model Setup

All simulations presented in this paper use the Global Atmosphere 7.2/Global Land 8.1 configuration (Walters et al., 2019) of the Met Office Unified Model (UM) at N768 resolution ($\simeq$ 25 km grid spacing near the equator) to drive a nested convection-permitting domain, centered around the Andenes ANX site. The limited area domain has a 0.01° by 0.01° horizontal grid spacing on a rotated grid ($\simeq$ 1 km grid spacing at the center of the domain), and covers an area of 600 x 600 grid points, using 90 stretched levels in the vertical, with a model top at an altitude of 40 km. The simulation domain and the location of the ANX site are shown in Figure 1. Simulations were run for days identified as cold-air outbreaks (CAO) (Geerts et al., 2022) during COMBLE (see Section 2.2). Simulations were initialized at 1200 UTC on the day preceding the CAO event and run for 36 hours. The first 12 hours were discarded as spin-up. Throughout the analysis, only CAO events are included.

The global model, initialized from the operational Met Office analysis, provided hourly lateral boundary data to the convection-permitting domain. Sea surface temperatures and sea ice coverage are obtained from the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA; Good et al. (2020)). The baseline convection-permitting configuration uses the Regional Atmosphere and Land 3 (RAL3) configuration at UM version 12.0. This configuration builds on the Regional Atmosphere 2 (RAL2) configuration described in Bush et al. (2022), using a blended turbulence boundary-layer parameterization (Boutle et al., 2014). The one-moment microphysics scheme with a single combined ice-snow category in RAL2 (Wilson & Ballard, 1999), was replaced in RAL3 by a new two-moment bulk microphysics scheme, CASIM (Cloud AeroSol Interacting Microphysics; Field et al. (2022)). CASIM predicts mass and number density of cloud water, rain, snow aggregates, ice crystals and graupel and uses in-cloud values of water content and number concentration in its collection calculations, using cloud fractions obtained from the macrophysics parameterization. Note that the parameterization of macrophysics (MacP) assumes a subgrid distribution of saturation-departure to determine the amount of condensation and the fractional cloud cover in the grid box. The microphysics parameterization (MicP) treats the conversion of this condensed liquid water content to other hydrometeor species and ultimately surface precipitation.

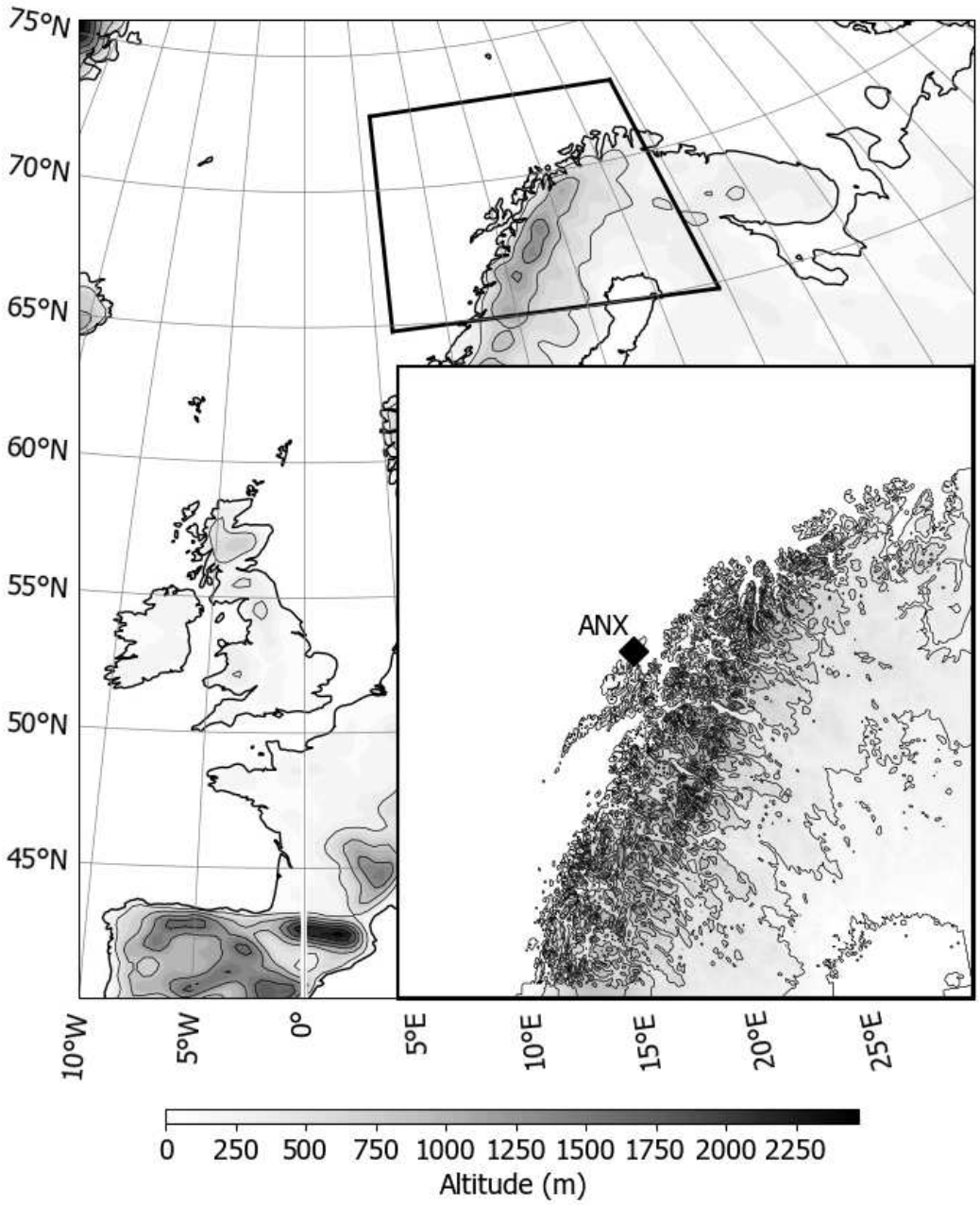

Figure 1. Location and spatial extent of the simulation domain used in all configurations. The location of the Andenes ANX supersite (69.14N, 15.68E) is indicated by the black diamond.

Note that two important modifications were made to CASIM compared to the default RAL3 configuration, with the goal to make the MicP and MacP assumptions more consistent. These modifications are summarized in Appendix A, and include more consistent treatment of the subgrid saturation-departure distribution and the subgrid phase-overlap of liquid and ice between the two parameterizations. These changes are indispensable to understand the relative role of micro- and macro-physics on mixed-phase cloud properties. The modifications in Appendix A are applied to all model configurations in the following analysis.

The SOCRATES scheme at version 21.07 (`https://code.metoffice.gov.uk/trac/socrates`, last access: 31 January 2022) (Edwards & Slingo, 1996; Manners et al., 2018) is used for the radiative transfer, using six SW and nine LW radiation bands. Radiative transfer of cloud droplets and ice crystals is treated following Edwards and Slingo (1996) and Baran et al. (2016) respectively. The radiation takes into account liquid and frozen CF calculated by the MacP, except for graupel, and additionally scales the liquid water content by a factor of 0.7 to represent in-cloud inhomogeneity in low clouds, following Cahalan et al. (1994). These radiation settings are identical to the RAL3 configuration. A 1- and 15-minute timestep was used for the dynamics and radiation respectively.

3.2 Model Experiments

An overview of all simulations is provided in Table 1. Simulations with a range of MacP setups are presented. We refer to the introduction of Van Weverberg, Morcrette, Boutle, Furtado, and Field (2021) for thorough discussion on MacPs and the role of the saturation-departure (SD) probability density function (PDF).

The baseline configuration builds on RAL3 using the bimodal macrophysics (BM) parameterization described in Van Weverberg, Morcrette, Boutle, Furtado, and Field (2021), treating liquid and ice cloud diagnostically. This parameterization derives cloud properties using the subgrid thermodynamic variability from the turbulence scheme, accounting for the co-existence of two modes of saturation-departure variability within a grid box near entrainment zones. As outlined in section 3.1 and Appendix A, the baseline configuration shown here differs from the default RAL3 configuration in its more consistent treatment of the subgrid saturation-departure distribution between the MicP and MacP, and the assumption of a minimum phase overlap between liquid and ice throughout the model. These modifications were implemented in all simulations listed below, unless stated otherwise (Appendix A).

A second configuration does not include a MacP, so grid boxes are either fully overcast or fully clear (NOMACP). A third experiment (SMITH) uses the diagnostic Smith (1990) scheme, assuming a triangular subgrid SD PDF with a fixed width, linked to a time- and space-invariant profile of critical relative humidity. Note that SMITH does not include the empirically adjusted CF, used in the RAL2 configuration (Van Weverberg, Morcrette, Boutle, Furtado, & Field, 2021) and that we passed on the Smith subgrid variability to CASIM, similar to BM and as explained in Appendix A.

Many observational studies (McFarquhar & coauthors, 2011; Abel et al., 2017; Zhang et al., 2019) find alternating single-phase pockets of ice and liquid in mixed-phase clouds. Abel et al. (2017) show improved q_{liq} in their mixed-phase cloud simulations when minimizing the overlap between liquid and ice. Note that the baseline configuration, BM, already assumes a minimum overlap in the MicP and MacP. An additional experiment, referred to as BM-MAX, is identical to BM, but assumes a maximum overlap between the liquid and ice phase. This is done by setting the overlap factor $\Phi = 1$ in Equation 1 in Appendix A. This experiment is aimed at understanding the importance of the overlap assumption relative to other modifications presented here.

Table 1. Experiment Overview

Experiment	Description
BM	Baseline Simulation using Regional Atmosphere and Land 3 Configuration, with Bimodal macrophysics (Van Weverberg, Morcrette, Boutle, Furtado, & Field, 2021) and CASIM microphysics (Field et al., 2022), including modifications listed in Appendix A for consistent treatment of saturation-departure between the macrophysics and the ice nucleation and deposition in the microphysics
$NOMACP$	As BM, but without a macrophysics parameterization
$SMITH$	As BM, but with Smith (1990) macrophysical parameterization
$BM - MAX$	As BM, but with a maximum phase overlap
HYB	As BM, but with the Hybrid macrophysics parameterization
$PC2$	As BM, but with the PC2 macrophysics parameterization
$HYB - RIM$	As HYB, but with only 1% of the snow riming rate in CASIM
$HYB - DEP$	As HYB, but with only 1% of the snow deposition rate in CASIM

Experiment descriptions and their acronyms used throughout this paper.

A next set of experiments revolves around the prognostic or diagnostic treatment of ice and liquid. BM and SMITH are fully diagnostic MacPs and treat frozen cloud fraction (CF_{fro}) very pragmatically. They obtain CF_{fro} by applying the diagnostic q_{liq}-CF_{liq} relation to the ice phase. (Note that ice and snow water contents are calculated in CASIM and (unlike for liquid) the role of the MacP is only to diagnose a CF_{fro}).

PC2 is a fully prognostic MacP and it uses CF tendencies from each part of the model that changes the thermodynamic state (Wilson et al., 2008). Hence, there are CF tendencies from the radiation, boundary-layer and microphysics scheme and clouds and condensate are advected. Furthermore, there are tendencies from the pressure change in the dynamics and an initiation/removal term if cloud ought to be initiated in a clear-sky grid box or reduced in an overcast grid box (Wilson et al., 2008).

Mostly, PC2 implicitly assumes symmetric, unimodal SD distributions, although a recent change includes the BM in its initiation/removal tendency. PC2 expects tendencies from CASIM as well, and this coupling between PC2 and CASIM was developed as part of this study (see Appendix B). Note that while CASIM includes two separate categories for ice and snow water *content*, it only assumes a single frozen *fraction* (CF_{fro}) for snow and ice combined. Each CASIM source term for cloud liquid or frozen condensate is assigned a CF tendency. This configuration (PC2) uses a minimum phase overlap for all tendency calculations as in BM.

Previous studies (Bush et al., 2020; Van Weverberg, Morcrette, & Boutle, 2021; Dipankar et al., 2020; Van Weverberg & Morcrette, 2022) indicate that diagnostic MacPs better capture mid-latitude low-level cloud and surface radiation, while high-level cloud and surface precipitation is better captured by PC2, especially in the tropics. Indeed, while the concept of instantaneous saturation adjustment for liquid clouds is well established, for ice clouds large supersaturations exist that invalidate this diagnostic framework.

Therefore, Van Weverberg and Morcrette (2022) propose a hybrid approach, representing liquid cloud diagnostically as in BM, but representing ice prognostically as in PC2. From Van Weverberg and Morcrette (2022), this hybrid scheme produces less compensating errors in simulations over the tropics, midlatitudes and the arctic than fully diagnostic MacPs. In this hybrid cloud scheme, implemented in experiment HYB, CF_{fro} is advected through the model and has microphysical tendencies as in PC2. Note that the CF_{fro} tendencies from CASIM were calculated as in PC2 (see Appendix B) and a minimum phase overlap was assumed.

Two last experiments have identical MacP settings as HYB, but focus on the microphysical snow production rates. As will be shown, the previous configurations predominantly produce large snow aggregates, while small ice crystals prevail in observations. HYB-RIM and HYB-DEP therefore assess the impact of reducing the CASIM snow riming growth and deposition respectively, by a factor of 100. While not viable options for operational configurations, these experiments target the role of the snow growth mechanisms in the overestimation of precipitation particle sizes and the depletion of q_{liq}.

4 Results

4.1 Role of Using a Macrophysics Parametrization

4.1.1 Vertically Integrated Model Biases

To explore the benefit of using a MacP in CAO simulations, this section focuses on the differences between NOMACP and all other simulations combined. Figure 2 provides an overview of the joint relative biases in downwelling SW radiation and 6 cloud-related variables, averaged over all cases. All biases have been reported for the STR (filled circles) and CNV (hollow circles) CAO modes separately as discussed in section 2.1. Further statistics are provided in Table 2, 3 and 4.

NOMACP has the most positive SW bias of all model experiments for both CAO modes (Figure 2). The SW bias is larger for the STR mode than for the CNV mode in all simulations, exceeding more than 70 % of the observed value in NOMACP. A substantial underestimation of the Liquid Water Path is present for both CAO modes and all simulations (LWP; Figure 2a). NOMACP has the largest LWP bias of all configurations (Table 2 and 4). The NOMACP Ice Water path (IWP) is underestimated in the STR mode, but is better captured in the CNV mode, hence the more modest SW bias in the latter (Table 2 and 4). Note that the simulated IWP in Table 2 includes all frozen hydrometeors, including ice, graupel and snow. All MacP configurations produce a similar shift in the IWP bias from the STR to the CNV mode, often flipping a negative bias into a positive bias (Figure 2b). Note that most IWP biases are within the observational uncertainty range. Low and high-level cloud cover is slightly overestimated by NOMACP (Figure 2c and d).

Table 3 reports two measures of stability, often used in shallow stratocumulus studies, as defined in Section 2.2. All simulations overestimate both the $\Delta_{SST-\Theta 850}$ and the EIS. While the former is a measure of the low-level instability, the latter provides an estimate of the inversion strength. In Figure 3, thermodynamic profiles are shown at ANX for all simulations. NOMACP, like most simulations, produces too cold and dry boundary layers, possibly related to the small LWP and reduced downwelling LW at this polar site. The EIS is defined as the contrast between the 700 hPa level and the surface. Most simulations have smaller biases near the 700 hPa level than at the surface, in particular for the CNV mode (Figure 3d), hence the generally stronger inversions. Mid-level latent heat release from vigorous glaciation could explain the difference between the CNV and STR mode biases.

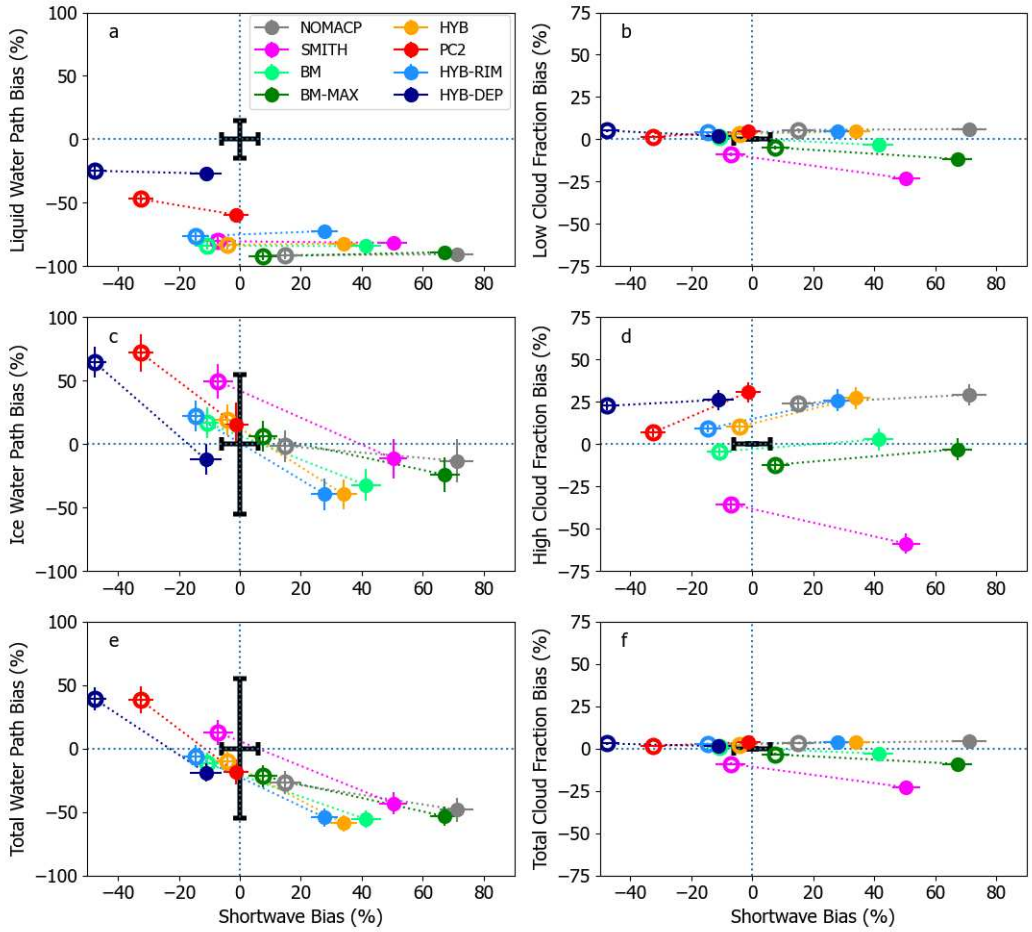

Figure 2. Joint downwelling shortwave (X-axis) and cloud property (Y-axis) relative biases for all model simulations, averaged over all CAO cases for the CNV (hollow circles) and STR (full circles) mode. Shown are (a) the liquid water path, (b) low cloud cover, (c) ice water path, (d) high cloud cover, (e) total water path, and (f) total cloud fraction. The boundary between low and high cloud is set to 3 km above ground level. Vertical and horizontal bars extending from the mean bias show the standard error of mean (SEM). The observational uncertainty, if known, is indicated by the black thick error bars extending from the origin. Observation sources include the ARM Best Estimate Cloud and Radiation Data product for radiation and the MicrobaseKaPlus for cloud properties. We refer to Section 2.2 for detailed information about these products and their processing. Dotted lines connect the CNV and STR bias for each model.

Table 2. Model evaluation statistics

Experiment	SW ($W\ m^{-2}$)	LW ($W\ m^{-2}$)	LWP ($kg\ m^{-2}$)	IWP ($kg\ m^{-2}$)	TWP ($kg\ m^{-2}$)	Low CF %	High CF %
OBS	78.2	276.3 (275.8)	0.150 (0.160)	0.379 (0.347)	0.528 (0.507)	94.8 (91.9)	69.0 (56.4)
Experiment	SW Bias ($W\ m^{-2}$)	LW Bias ($W\ m^{-2}$)	LWP Bias ($kg\ m^{-2}$)	IWP Bias ($kg\ m^{-2}$)	TWP Bias ($kg\ m^{-2}$)	Low CF Bias %	High CF Bias %
NOMACP	11.7	-3.6 (*-1.1*)	-0.136 (-0.144)	*-0.006* (0.060)	-0.142 (-0.084)	5.2 (**8.1**)	16.6 (**20.8**)
SMITH	-5.6	6.2 (7.5)	-0.120 (-0.131)	0.187 (0.127)	0.067 (*-0.004*)	**-8.4** (-6.6)	**-24.5** (-19.6)
BM	-8.5	11.9 (14.1)	-0.125 (-0.135)	0.063 (0.049)	-0.062 (-0.085)	*1.3* (3.0)	*-3.1* (*-3.4*)
BM − MAX	6.0	*3.2* (6.5)	**-0.138** (**-0.145**)	0.024 (*0.038*)	-0.114 (-0.108)	-4.6 (*-2.8*)	-8.5 (-7.3)
HYB	*-3.3*	5.7 (8.3)	-0.124 (-0.132)	0.072 (0.050)	-0.052 (-0.083)	3.1 (6.3)	7.0 (6.7)
PC2	-25.4	9.9 (11.4)	-0.070 (-0.084)	**0.272** (0.178)	0.202 (0.093)	1.5 (4.4)	4.7 (7.5)
HYB − RIM	-11.3	7.3 (9.0)	-0.114 (-0.119)	0.083 (0.063)	*-0.032* (-0.055)	3.8 (6.5)	6.3 (6.3)
HYB − DEP	**-37.3**	**14.7** (**16.5**)	*-0.037* (*-0.032*)	0.244 (**0.231**)	**0.207** (**0.199**)	4.9 (7.3)	15.5 (18.5)
Experiment	SW RMSE ($W\ m^{-2}$)	LW RMSE ($W\ m^{-2}$)	LWP RMSE ($kg\ m^{-2}$)	IWP RMSE ($kg\ m^{-2}$	TWP RMSE ($kg\ m^{-2}$)	Low CF RMSE %	High CF RMSE %
NOMACP	**106.7**	**37.2** (**36.6**)	**0.188** (**0.216**)	1.216 (1.140)	1.231 (1.148)	19.7 (25.1)	49.4 (**54.2**)
SMITH	98.5	30.8 (31.8)	0.181 (0.204)	1.344 (1.143)	1.343 (1.151)	**31.2** (**33.6**)	**55.7** (52.4)
BM	*81.8*	28.5 (31.4)	0.175 (0.203)	1.179 (*1.020*)	1.171 (*1.028*)	21.4 (26.6)	45.3 (48.1)
BM − MAX	93.6	30.0 (30.2)	0.183 (0.209)	*1.156* (1.053)	*1.156* (1.065)	25.0 (28.2)	50.9 (50.6)
HYB	88.2	27.2 (29.2)	0.179 (0.203)	1.166 (1.066)	1.176 (1.081)	20.5 (25.4)	44.8 (48.1)
PC2	87.3	*26.4* (28.8)	*0.142* (*0.173*)	**1.459** (**1.231**)	**1.457** (**1.246**)	21.2 (26.0)	44.1 (48.2)
HYB − RIM	90.2	26.5 (*28.5*)	0.171 (0.192)	1.174 (1.081)	1.181 (1.093)	20.3 (*25.0*)	*43.7* (*47.9*)
HYB − DEP	89.4	27.0 (29.4)	0.165 (0.193)	1.229 (1.099)	1.257 (1.141)	*19.7* (25.1)	45.9 (50.6)

Overview of model evaluation statistics for the CNV CAO mode at the ANX supersite. Provided are (first row) the time-averaged observed value of surface downwelling shortwave radiation (SW), surface downwelling longwave radiation (LW), liquid water path (LWP), ice water path (IWP), total water path (TWP), low-level cloud fraction (Low CF) and high-level cloud fraction (High CF). The boundary between low and high cloud is set to 3 km above ground level. Data provided is for times with zenith angles smaller than 90° only, although statistics for the full dataset are provided in brackets. Observation sources include the ARM Best Estimate Cloud and Radiation Data product for radiation and the MicrobaseKaPlus for cloud properties. We refer to Section 2.2 for detailed information about these products and their processing. Furthermore, the absolute time-averaged bias and root-mean squared error (RMSE) between each of the model simulations and the observations are provided. The largest and smallest bias magnitude for each variable are highlighted in boldface and italic respectively.

Table 3. Stability evaluation statistics

Experiment	CNV EIS K	CNV $\Delta_{SST-\Theta 850}$ K	STR EIS K	STR $\Delta_{SST-\Theta 850}$ K
OBS	-17.4 (-17.2)	0.5 (0.5)	-19.1 (-18.0)	4.1 (4.0)

Experiment	CNV EIS Bias K	STR $\Delta_{SST-\Theta 850}$ Bias K	STR EIS Bias K	STR $\Delta_{SST-\Theta 850}$ Bias K
NOMACP	3.2 (2.4)	**0.8** (**0.9**)	4.0 (3.8)	0.5 (0.7)
SMITH	3.9 (3.2)	0.7 (0.8)	4.8 (4.4)	0.4 (0.6)
BM	4.1 (3.4)	0.6 (0.7)	**5.0** (**4.7**)	0.4 (0.5)
BM − MAX	**4.2** (**3.5**)	0.7 (0.7)	5.0 (4.6)	0.4 (0.5)
HYB	2.7 (1.9)	0.7 (0.8)	4.1 (3.9)	0.6 (0.6)
PC2	2.8 (2.1)	*0.5* (*0.6*)	4.2 (3.9)	*0.3* (*0.4*)
HYB − RIM	2.7 (2.0)	0.8 (0.8)	4.2 (3.9)	**0.6** (**0.7**)
HYB − DEP	*1.2* (*0.8*)	0.8 (0.8)	*3.4* (*3.1*)	0.4 (0.5)

Overview of model evaluation statistics of two stability measures for the STR and CNV CAO mode at the ANX supersite. Provided are the Estimated Inversion Strength (EIS), and $\Delta_{SST-\Theta 850}$. Observed values are calculated from the interpolated sounding product. We refer to Section 2.2 for detailed information about these products and their processing. The absolute time-averaged bias is provided for all simulations and the largest and smallest bias magnitude for each variable are highlighted in boldface and italic respectively.

Simulations presented here use the OSTIA SSTs, which are generated by ingesting observations. The contrast between the SST and the cold bias near the 850 hPa level increases instability in the lower troposphere and hence the ($\Delta_{SST-\Theta 850}$; Figure 3d). The cold boundary-layer bias over a relatively warm sea could invigorate updrafts and enhance glaciation, which would then further deplete liquid water and possibly further cool the boundary layer. It is worth exploring this possible feedback mechanism in future studies.

4.1.2 Cloud regime analysis

To disentangle which of the cloud biases matter most from a cloud-radiative perspective, a detailed cloud-radiation regime analysis is performed. Cloud optical thickness is affected by cloud fraction, cloud phase, cloud water content and cloud particle scattering properties (e.g. size and shape). Data on biases in the first three of these determining factors are available each 15 minutes for all simulations. Each data point can be assigned to a negative, neutral or positive relative bias in the cloud fraction (CF), liquid fraction (LF; defined as $q_{liq}/(q_{liq}+q_{fro})$) and total water path ($TWP = IWP + LWP$), respectively, yielding 27 possible combined cloud property bias regimes. Negative and positive cloud property biases are defined as respectively an underestimation and overestimation by at least 50%. The average radiation bias can be calculated for each of these 27 regimes. Multiplication of these radiation biases with their relative regime frequencies yields the contribution of each regime to the total radiation bias. The total radiation bias consists of the sum of all 27 radiation biases multiplied by their respective regime frequencies.

Figure 4 shows a decomposition of the SW (left) and LW (right) downwelling bias into the cloud property bias regimes, for the all CNV (top) and STR (bottom) CAO events during COMBLE. Note that we first split between the regimes having a negative and

Table 4. Model evaluation statistics

Experiment	SW ($W\ m^{-2}$)	LW ($W\ m^{-2}$)	LWP ($kg\ m^{-2}$)	IWP ($kg\ m^{-2}$)	TWP ($kg\ m^{-2}$)	Low CF %	High CF %
OBS	119.3	298.2 (298.2)	0.190 (0.237)	0.231 (0.250)	0.420 (0.487)	94.2 (91.7)	35.8 (39.5)
Experiment	SW Bias ($W\ m^{-2}$)	LW Bias ($W\ m^{-2}$)	LWP Bias ($kg\ m^{-2}$)	IWP Bias ($kg\ m^{-2}$)	TWP Bias ($kg\ m^{-2}$)	Low CF Bias %	High CF Bias %
NOMACP	**84.8**	**-21.5** (**-20.9**)	**-0.172** (**-0.219**)	-0.031 (-0.037)	-0.202 (-0.256)	5.8 (7.9)	10.4 (12.1)
SMITH	60.0	-11.7 (-10.9)	-0.155 (-0.201)	-0.027 (-0.039)	-0.181 (-0.240)	**-21.8** (**-19.3**)	**-21.1** (**-19.6**)
BM	49.3	-3.7 (-4.2)	-0.160 (-0.204)	-0.075 (-0.104)	-0.234 (**-0.308**)	-3.2 (*-0.9*)	*1.0* (5.6)
BM − MAX	80.1	-14.8 (-12.3)	-0.169 (-0.213)	-0.056 (-0.058)	-0.224 (-0.270)	-10.9 (-6.5)	-1.1 (*4.8*)
HYB	40.6	-3.0 (-2.3)	-0.155 (-0.199)	**-0.092** (**-0.107**)	**-0.247** (-0.306)	4.4 (6.4)	9.7 (14.3)
*PC*2	*-1.7*	4.9 (5.2)	-0.113 (-0.153)	0.035 (0.024)	*-0.077* (-0.129)	4.6 (6.6)	10.9 (15.1)
HYB − RIM	33.1	*-0.6* (*0.2*)	-0.137 (-0.176)	-0.091 (-0.099)	-0.228 (-0.275)	4.6 (6.6)	9.2 (13.8)
HYB − DEP	-13.3	3.9 (4.1)	*-0.051* (*-0.092*)	*-0.028* (*-0.000*)	-0.079 (*-0.093*)	*1.9* (4.4)	9.3 (14.7)
Experiment	SW RMSE ($W\ m^{-2}$)	LW RMSE ($W\ m^{-2}$)	LWP RMSE ($kg\ m^{-2}$)	IWP RMSE ($kg\ m^{-2}$	TWP RMSE ($kg\ m^{-2}$)	Low CF RMSE %	High CF RMSE %
NOMACP	**169.3**	**40.7** (**40.1**)	**0.190** (**0.272**)	**0.864** (**0.856**)	**0.889** (**0.912**)	21.1 (25.8)	51.3 (50.5)
SMITH	132.8	29.4 (28.8)	0.176 (0.257)	0.774 (0.772)	0.804 (0.835)	**38.5** (**38.2**)	50.3 (*47.1*)
BM	136.9	27.4 (26.7)	0.179 (0.259)	0.630 (*0.622*)	0.669 (*0.709*)	26.1 (27.8)	50.8 (49.7)
BM − MAX	148.6	30.7 (29.6)	0.186 (0.266)	0.696 (0.731)	0.733 (0.796)	30.6 (31.0)	51.6 (50.4)
HYB	125.2	25.8 (26.1)	0.177 (0.257)	0.614 (0.639)	*0.656* (0.721)	21.3 (*25.3*)	51.8 (**51.3**)
*PC*2	*106.8*	*22.5* (*23.6*)	*0.148* (*0.228*)	0.858 (0.856)	0.884 (0.902)	*21.2* (25.6)	48.0 (49.3)
HYB − RIM	120.9	23.4 (24.0)	0.176 (0.258)	0.635 (0.711)	0.676 (0.776)	21.5 (25.7)	**52.0** (51.1)
HYB − DEP	134.5	25.0 (24.2)	0.172 (0.235)	*0.614* (0.710)	0.675 (0.790)	23.3 (25.8)	*47.1* (48.6)

As Table 2, but for the STR CAO mode.

manuscript submitted to *JGR: Atmospheres*

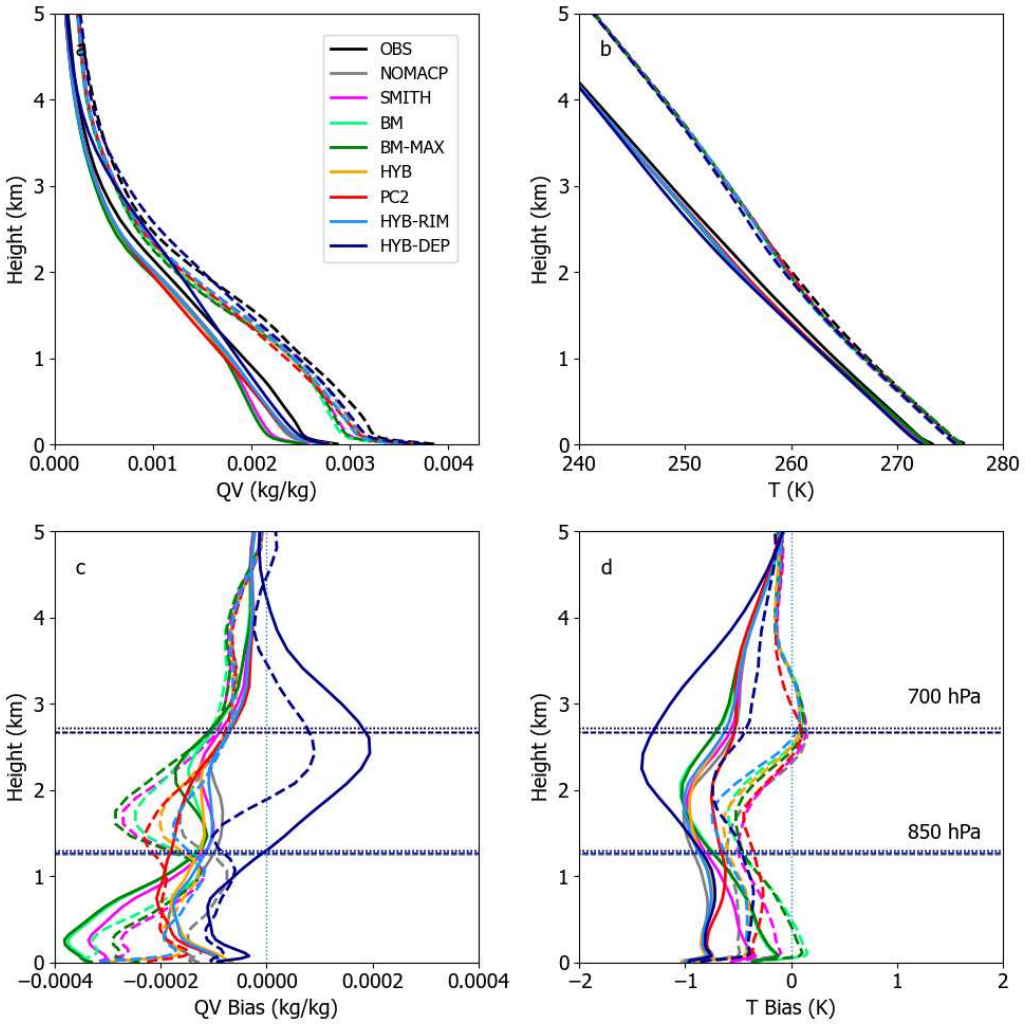

Figure 3. Vertical profiles of (a) specific humidity (q_v) and (b) temperature (T) and (c) and (d) their bias against the interpolated sounding data for the location of the ANX supersite, averaged over all CNV (dashed lines) and STR (solid lines) CAO cases during COMBLE. The horizontal dotted lines in panel c and d show the mean altitude of the 700 hPa (top) and 850 hPa levels, used in the calculation of the stability metrics discussed in the text.

positive radiation bias, and then stack each of their radiation contributions. A combination of three hatching/shading patterns describes which cloud property bias combinations explain most of the total radiation bias, as outlined in the legend above the figure. From this figure, it becomes immediately clear that the WP bias determines the radiation biases to a large extent in all simulations. Indeed, positive SW biases (bars above the zero line in the left hand panels) are predominantly associated with negative WP biases (blue colors), and vice versa (red bars below the zero line). Furthermore, negative LF biases (indicated by loose stippling) prevail, regardless of the sign of the SW and LW biases. However, there are notable differences between the different simulations that will be explored in the next sections.

The positive SW bias in the NOMACP CNV mode hides a partial bias compensation. Times when the SW bias is positive prevail and are associated with mostly too small WP (blue bars), as well as too small LF (loose stippling). The CF in NOMACP is well captured and not responsible for radiation biases (no diagonal or crossed hatching). Times with negative SW biases exist as well, and are always due to excessive WP (red bars), while these clouds still lack liquid (loose stippling).

The SW biases in the NOMACP STR mode are more decidedly positive with little bias compensation. The radiation biases in this case are predominantly explained by the cloud phase (loose stippling), although still half of the SW bias is related with a significant underestimation of the WP (blue bars).

4.1.3 Vertical Profiles

Unfortunately, no aircraft measurements were collected during COMBLE. However, two phase-identification algorithms were applied using ground-based remote sensing measurements that could provide insight in vertical phase distribution, as described in Section 2.2. While MBase identifies the phase mainly using temperature thresholds and is therefore less reliable in its detection of supercooled q_{liq} at cold temperatures, TPhase uses Doppler velocities and spectra in its phase detection (Section 2.2). Liquid, mixed-phase and ice clouds in both retrievals and the simulations are defined consistently as having $CF_{liq} > 0.1$, $CF_{liq} > 0.1$ and $CF_{fro} > 0.1$, and $CF_{fro} > 0.1$ respectively.

Figure 5 shows the frequency of liquid, ice and mixed-phase in the simulations and the two retrievals. Note that the observed profiles were constructed using the time-averaging technique outlined in section 2.2. The grey shading surrounding the TPhase and MBase profiles in Figure 5 indicates the sensitivity to doubling and halving the environmental wind speeds in the time averaging method. There are some notable differences between the CNV and STR modes in the observations. The CNV mode has more mixed-phase cloud at mid-levels than the STR mode, at least according to TPhase (Figure 5a-d). Close to the surface, ice is prevalent in TPhase, while mixed-phase is predominantly identified by MBase. Figure 5 confirms the lack of mixed-phase and liquid in NOMACP, and its excessive glaciation, in particular in the lowest 2 km of the atmosphere (Figure 5c and g).

Figure 6 shows the frequency of small ice and snow aggregates separately. Note that TPhase and the simulations employ different definitions of the boundary between small and large ice, and hence caution is needed when directly comparing the model and the observations. Nevertheless, all configurations hardly produce any ice crystals and transfer all frozen condensate into the large snow aggregate category. Even with the caveat of inconsistent definitions, this difference is remarkable. Additional evidence for too large precipitation particles will be given in the next sections. From Figure 6b, NOMACP has by far the largest frequency of large snow aggregates near the surface.

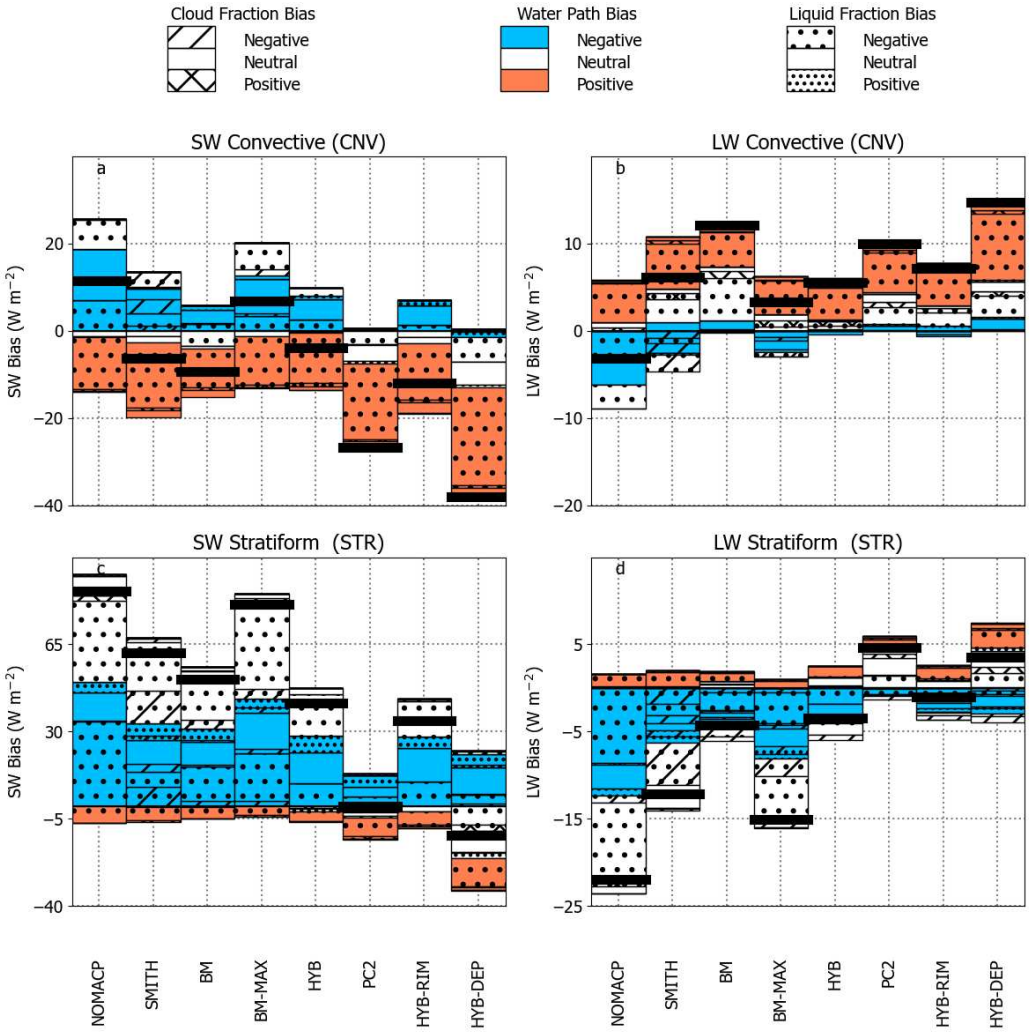

Figure 4. Cloud regime analysis for all simulations, averaged over all CNV (top) and STR (bottom) CAO events during COMBLE. Cloud regimes are based on negative, neutral and positive biases in CF, TWP, and LF respectively (using a relative bias of ± 50% to discriminate between the bias regimes). More detailed definitions of the regimes are provided in the text. Panels on the left and right show the decomposition of respectively the average surface downwelling shortwave (SW) and longwave (LW) bias in cloud-bias regimes, using three different hatching patterns as outlined in the legend. The bold black horizontal lines overlaid over each of the bias decompositions denotes the total radiation bias for each simulation.

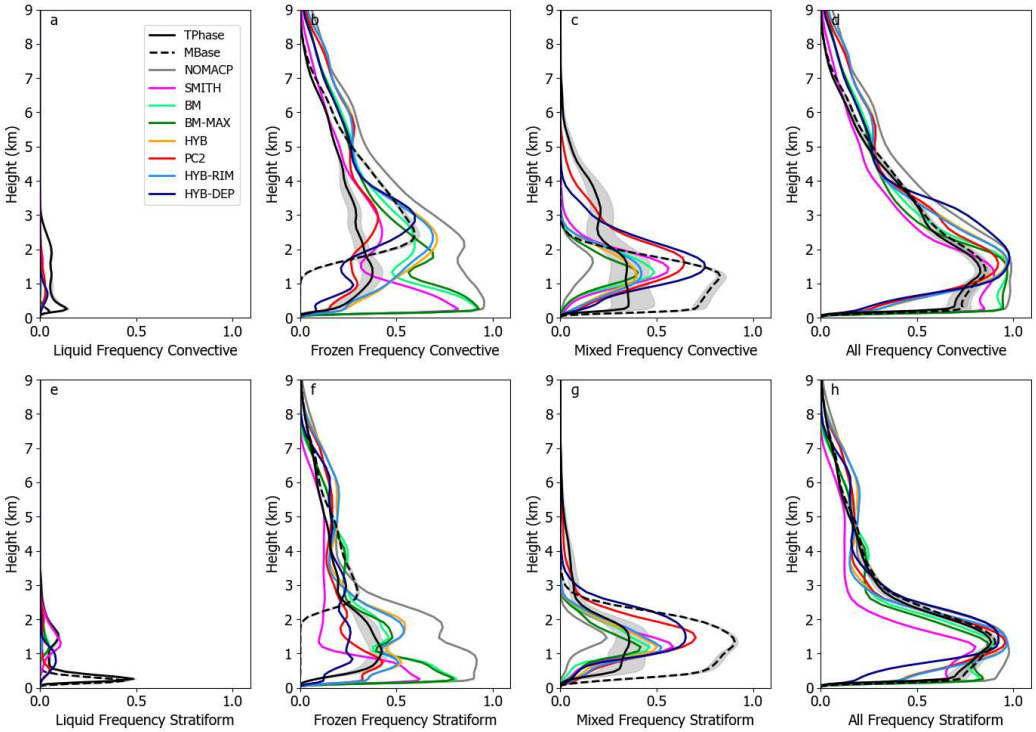

Figure 5. Time-average vertical profiles of cloud phase as observed by MBase and TPhase and as simulated by all configurations for the CNV (a, b, c, d) and STR (e, f, g, h) CAO events during COMBLE. Note that TPhase profiles are provided with and without temporal averaging to provide an estimate of the uncertainty. As outlined in the text, if the fraction of ice- (liquid-) containing hydrometeors is larger than 0.1 in the model grid box, or the observed time-window representative of a model grid length, the cloud is defined as a frozen (liquid) cloud. If both the ice and liquid-containing hydrometeor fraction are larger than 0.1, the cloud is assumed to be mixed-phase.

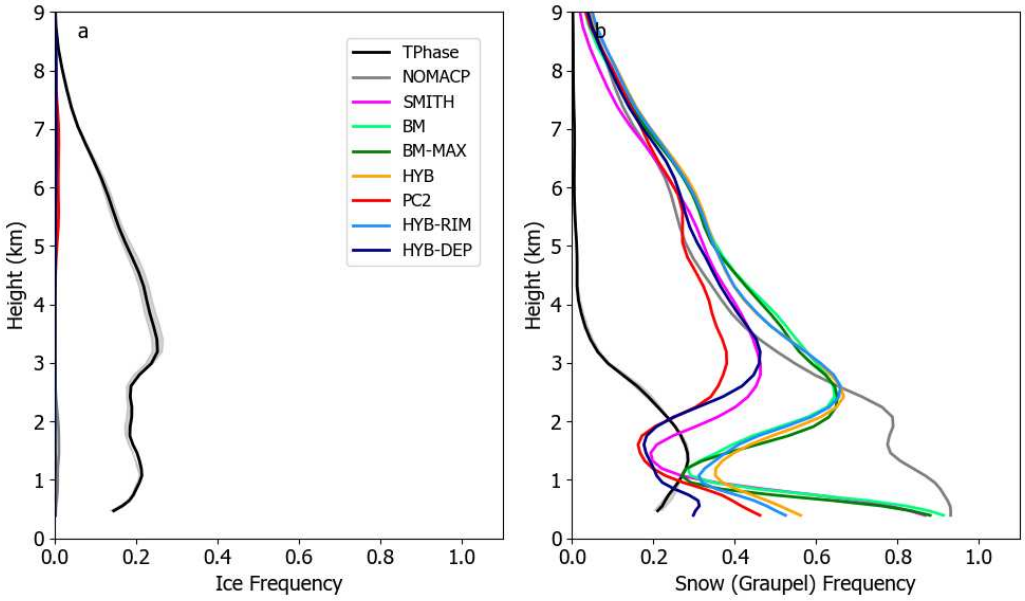

Figure 6. Time-average vertical profiles of (a) Ice and (b) Snow frequencies observed by TPhase and simulated by all configurations for all CAO events during COMBLE. Note that the ice and snow definitions between the observations and the simulations are not entirely consistent.

4.1.4 Surface Precipitation

Given the excessive glaciation in NOMACP, it is worth exploring the implications for surface precipitation. Statistics using two observational sources of surface precipitation (PWD and WB) are shown Table 5. Note that due to the frequent mixed-phase precipitation at ANX, quantitative precipitation retrievals from radar (i.e. conventional empirical Z-R relationships), disdrometer or optical rain gauges are often unreliable. Rain gauge data will also be associated with considerable uncertainty in these mixed-phase conditions. All simulations replicate the heavier and more frequent precipitation in the CNV CAO mode and produce precipitation rates within the observational uncertainty for both modes. NOMACP has slightly heavier precipitation than many other simulations in the CNV mode, consistent with its more excessive glaciation (Table 5).

4.2 Role of Subgrid PDF in diagnostic MacPs

4.2.1 Vertically Integrated Model Biases

Ignoring subgrid cloud variability (NOMACP) leads to a lack of supercooled liquid, excessive glaciation, and large radiation biases. However, from Figure 2, the use of a MacP does not automatically lead to improved cloud properties. The diagnostic MacPs (BM and SMITH) still exhibit considerable radiation biases, in particular in the STR mode.

This section concentrates on the difference between BM and SMITH and the role of their subgrid SD PDF. As mentioned in Section 3, SMITH assumes a unimodal, symmetric and triangular subgrid SD PDF with a fixed variance. BM assumes a combination of two Gaussian PDFs with variance linked to the turbulent properties, allowing for skewed mixture distributions (Van Weverberg, Morcrette, Boutle, Furtado, & Field, 2021).

Table 5. Surface precipitation evaluation statistics

Experiment	Mean CNV (mm day^{-1})	STD CNV (mm day^{-1})	Mean STR (mm day^{-1})	STD STR (mm day^{-1})
PWD	1.61 ± 0.16	7.17	0.49 ± 0.05	3.74
WB	0.92 ± 0.55	6.44	0.11 ± 0.06	2.04
$NOMACP$	1.37	7.94	0.25	2.44
$SMITH$	1.38	6.63	0.33	2.47
BM	1.28	6.08	0.31	2.64
$BM-MAX$	1.34	5.98	0.34	3.00
HYB	1.26	7.76	0.27	2.79
$PC2$	1.07	6.37	0.28	2.91
$HYB-RIM$	1.28	7.74	0.26	2.66
$HYB-DEP$	0.82	4.99	0.17	2.02

Surface Precipitation statistics for (left) the CNV and (right) the STR CAO mode at the location of the ANX supersite for the observations and all simulations. Provided are the time-averaged precipitation rate (Mean) and the Standard Deviation of the hourly precipitation rate (STD). Two observational sources are shown: the present weather detector (PWD) and the weighing bucket pluviometer (WB). Observational uncertainty for PWD and WB is indicated with the ± symbol, indicating the confidence interval

Both schemes exhibit a similarly negative LWP bias (Figure 2a). From Figure 2 and Tables 2 and 4, the radiation bias in BM is only slightly smaller than in SMITH. CF in SMITH is underestimated but better captured in BM (Figure 2b, d and f). Conversely, the IWP is underestimated in BM and overestimated in SMITH (Figure 2a and c). The next section discusses which of these contrasting cloud biases contributes most to the radiation bias differences between BM and SMITH.

4.2.2 Cloud regime analysis

Cloud regime analysis for the diagnostic MacPs is shown in Figure 4. In the CNV mode, the mean SW bias magnitude is smaller in SMITH than in BM, although there is more error compensation (Figure 4a). Compared to BM, SMITH exhibits more frequent positive SW biases (particularly when CF is biased low; diagonal hatching). Conversely, SMITH also experiences more frequent negative SW biases associated with overestimating the WP (but still underestimating the LF).

SW radiation in both SMITH and BM is underestimated in the CNV mode, despite underestimated LWP and/or CF (although IWP is overestimated in SMITH). This might point to further errors in the cloud radiative properties (e.g. the effective radii or scattering properties in the radiation parameterization) that should be investigated in future studies.

In the STR mode, less bias compensation is present and SW radiation is persistently overestimated (Figure 4c). In both SMITH and BM, this predominantly related to underestimating the LF (loose stippling), although at least half the bias is also associated with too small WP. The positive SW bias is larger in SMITH, which can be linked to more frequent CF underestimations (diagonal hatching), consistent with Figure 2b, d and f. Compared to NOMACP both diagnostic schemes have smaller contributions to the radiation biases from LF biases, indicating the improved representation of mixed-phase conditions from using a MacP.

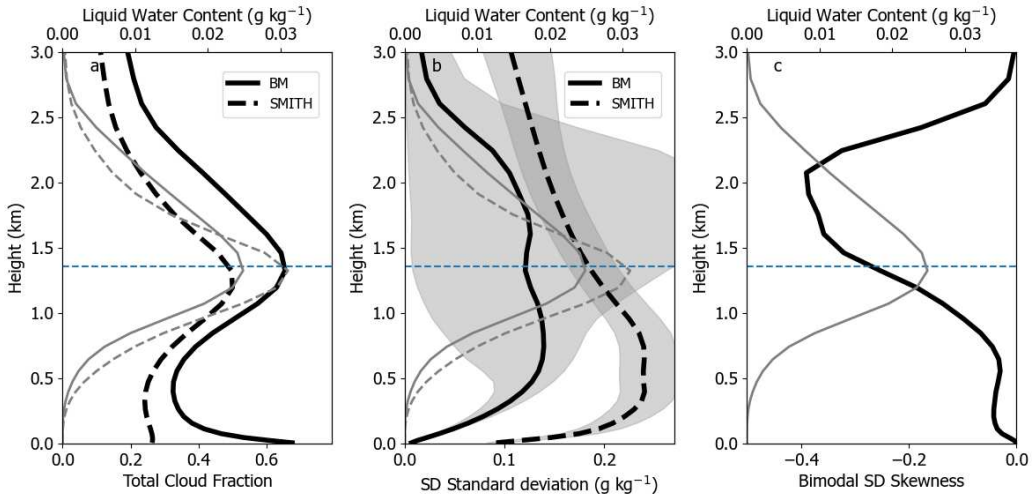

Figure 7. Vertical profiles of cloud properties in BM and SMITH, averaged over all STR CAO cases during COMBLE. Panel (a) shows the average total cloud fraction (black) for BM (solid) and SMITH (dashed). Panel (b) shows the average saturation-departure standard deviation (black) in BM (solid) and as obtained from the critical relative humidity profile in SMITH (dashed). The shaded areas denote the 10^{th} and 90^{th} percentiles of the standard deviations. Panel (c) shows the average saturation-departure skewness associated with the BM. The Liquid Water Content (q_{liq}) for BM (solid) and SMITH (dashed) are overlayed in each plot as the grey lines. The dotted horizontal line shows the average boundary layer height.

4.2.3 Vertical Profiles

The two diagnostic MacPs produce less frequent pure frozen and more frequent mixed-phase conditions than the NOMACP in both the CNV and STR mode (Figure 5), in particular between 1 and 2 km above ground level. Mid-level CNV and STR clouds and near-surface STR clouds are still too glaciated in SMITH and BM, while the frequency of mixed-phase near the boundary-layer top appears better captured compared to TPhase. Near the surface in the STR mode, observed clouds are often identified to be pure liquid, while pure frozen condensate dominates the lower levels in the diagnostic MacPs (note again the higher confidence in TPhase than in MBase for observed vertical phase distribution).

To explore the reasons for the smaller CF in SMITH compared to BM, causing more positive SW biases in the former (see previous sections), Figure 7 shows vertical profiles of the macrophysical cloud properties of both schemes. The SD variance in SMITH can be obtained from the time- and space- invariant critical relative humidity profile (and assuming a triangular SD PDF). In contrast to midlatitude regions (Van Weverberg, Morcrette, & Boutle, 2021), the SMITH variance in the arctic is larger than the turbulence-based BM variance (Figure 7b), even near the entrainment zone. In a unimodal PDF, larger variances are associated with larger q_{liq}, but also more partial cloudiness. The large SD variance in SMITH prevents cloud from becoming overcast (Figure 7a). Linking the underlying SD distribution to turbulent properties, as in BM, gives more realistic behavior of large variability in mid-latitude stratocumulus, but more quiescent conditions in polar regions.

It is also noteworthy that SMITH has a smaller overall condensate frequency than BM (Figure 5d and h), but large TWP (Figure 2e). In combination with its smaller CF, this increases in-cloud water paths, with important implications for the microphysical

Table 6. Ice-phase production rates

Experiment	Riming % ($10^8 kgkg^{-1}s^{-1}$)	Deposition % ($10^8 kgkg^{-1}s^{-1}$)	Nucleation % ($10^8 kgkg^{-1}s^{-1}$)
$NOMACP$	15.6 (0.21)	84.4 (1.16)	0.0 (0.00)
$SMITH$	34.4 (0.51)	65.6 (0.98)	0.0 (0.00)
BM	10.2 (0.13)	89.8 (1.12)	0.0 (0.00)
$BM-MAX$	12.5 (0.16)	87.5 (1.15)	0.0 (0.00)
HYB	10.5 (0.15)	89.5 (1.32)	0.0 (0.00)
$PC2$	39.3 (0.83)	60.5 (1.28)	0.2 (0.00)
$HYB-RIM$	0.2 (0.00)	99.7 (1.49)	0.0 (0.00)
$HYB-DEP$	31.1 (0.40)	68.9 (0.88)	0.0 (0.00)

Fraction of total microphysical production, and average production rate (parentheses) of the ice phase from riming, deposition and nucleation for all experiments. Riming and deposition rates for ice, snow and graupel are combined in the Table.

conversion rates. Indeed, from the relative importance of riming, deposition and nucleation in the frozen-condensate production rates, it is clear that SMITH is much more riming-dominated compared to BM.

4.2.4 Role of Phase Overlap

A next sensitivity experiment involves the role of the phase overlap. While MacPs predict the CF_{liq} and CF_{fro}, they usually make fairly crude assumptions on how these phases overlap within a grid box. This can potentially have important repercussions as observations suggest that mixed-phase clouds often show alternating pockets of single-phase liquid or frozen condensate. This phase separation can put breaks on the glaciation as less liquid is readily available for consumption by ice through riming or the Bergeron process.

As outlined in section 3, BM-MAX assumes a maximum phase overlap (hence minimizing the amount of mixed-phase), in contrast to the minimal phase overlap in BM (Appendix A). From Figure 2 and Table 2 and 4, SW radiation is considerably degraded in BM-MAX. This is due to a further reduction in LWP (most obvious in the CNV mode) and a negative bias in CF. As expected, this leads to a reduction of mixed-phase conditions (Figure 5c and g) compared to BM. The SW bias is more positive in BM-MAX than in BM, which is mainly associated with a more frequent bias in LF, but - in particular in the CNV mode - also with more important underestimations of the WP. The enhanced glaciation in BM-MAX also leads to slightly larger precipitation rates (Table 5), consistent with findings of Abel et al. (2017).

4.3 Role of Diagnostic versus Prognostic MacP Approach

4.3.1 Vertically Integrated Model Biases

Different approaches exist to represent subgrid CF in numerical models, as outlined in section 3. BM is fully diagnostic and calculates q_{liq}, CF_{liq} and CF_{fro} at each time step from the large-scale state, without memory of previous timesteps. Conversely, the prognostic PC2 predicts CF tendencies of the liquid and frozen cloud phase (Wilson et al., 2008). A newly designed hybrid approach, implemented in HYB, combines prognostic CF_{fro} as in PC2 with diagnostic CF_{liq} as in BM (Section 3 and Van Weverberg and Morcrette (2022)).

Note that BM, HYB and PC2 all include the improved coupling between the MicP and MacP as outlined in Appendix A. Of all the sensitivities shown so far, the inclusion of memory has the most pronounced impact on q_{liq} (Figure 2a). For the CNV mode, HYB has the most modest SW bias, while PC2 exhibits a substantial negative SW bias. A large positive SW bias exists in the STR mode with BM and HYB, while in this case, PC2 performs much better.(Figure 2 and Table 2 and 4). PC2 has significantly larger LWP and IWP compared to HYB and BM (Figure 2a and c).

4.3.2 Cloud regime analysis

The cloud radiation regime analysis yields fairly similar results for BM and HYB, although the SW bias is slightly improved in both the CNV and STR modes in the HYB configuration (Figure 4a and c). PC2 on the other hand has an important negative SW bias in the CNV mode, associated with too large WP, even if the clouds still lack liquid (red shading and loose stippling in Figure 4a). For the STR mode, PC2 has the smallest SW bias and by far the least bias compensation of all configurations (Figure 4c).

4.3.3 Vertical Profiles

From Figure 5b and f, a prognostic representation of the frozen CF (as in HYB and PC2) drastically changes the pure frozen hydrometeor frequencies in the lower atmosphere, compared to the diagnostic schemes. Indeed, there is relatively less frozen condensate (and indeed, total condensate) near the surface, and relatively more condensate near the boundary-layer top. In PC2, much more frequent mixed-phase conditions are present at all levels (Figure 5c and g). At least for the mid-levels of the atmosphere the latter is in better agreement with the benchmark observation TPhase.

Figure 8 shows joint probability density functions of q_{liq} and q_{fro} and CF as observed in MBase, and as simulated by BM, HYB and PC2. As noted in Section 2.2, confidence in MBase for the liquid versus frozen condensate partitioning is rather small for polar clouds, although the vertically distributed cloud mask and the liquid water paths should be more reliable (which are constrained by radar reflectivities and the ground-based microwave radiometers respectively). This is also evident from Figure 5 showing a very limited height range for the occurrence of liquid in MBase compared to TPhase, related to the temperature-dependency in the latter.

Hence, while the observed CF distributions in Figure 8 are fairly reliable, comparison with the ice and liquid water contents between the model and the observations requires serious caution. We will therefore focus on the differences between the model simulations rather than on an evaluation against observations in the following analysis.

BM produces a tight relation between the q_{liq} and CF_{liq} and q_{fro} and CF_{fro} (Figure 8b and f). Furthermore, liquid clouds mostly have partial cloud cover, and even ice clouds are frequently broken (histograms to the right of the panels in Figure 8). While there is considerable uncertainty in the partitioning of ice and liquid in MBase, there is more confidence in the typically near-overcast observed conditions (Figure 8a and e). Given the relatively small variances associated with the BM in this polar environment (Figure 7), it is remarkable that liquid clouds struggle to reach overcast conditions. It is possible that the competition for water vapour from the ice phase, as formulated following Furtado et al. (2016) prevents liquid cloud from reaching larger cloud cover. Frozen CF in BM is diagnosed using the q_{liq}-CF_{liq} relation and this clearly leads to the same tight coupling between q_{fro} and CF_{fro} (Figure 8f), although full frozen cloud cover is often achieved. The q_{liq} distribution is weighted to too small values (bottom histogram in Figure 8c; note that these observations are constrained by the microwave radiometer LWP measurements (Section 2.2). The q_{fro} appears well-captured (Figure 8f), although note that its observed values are less reliable.

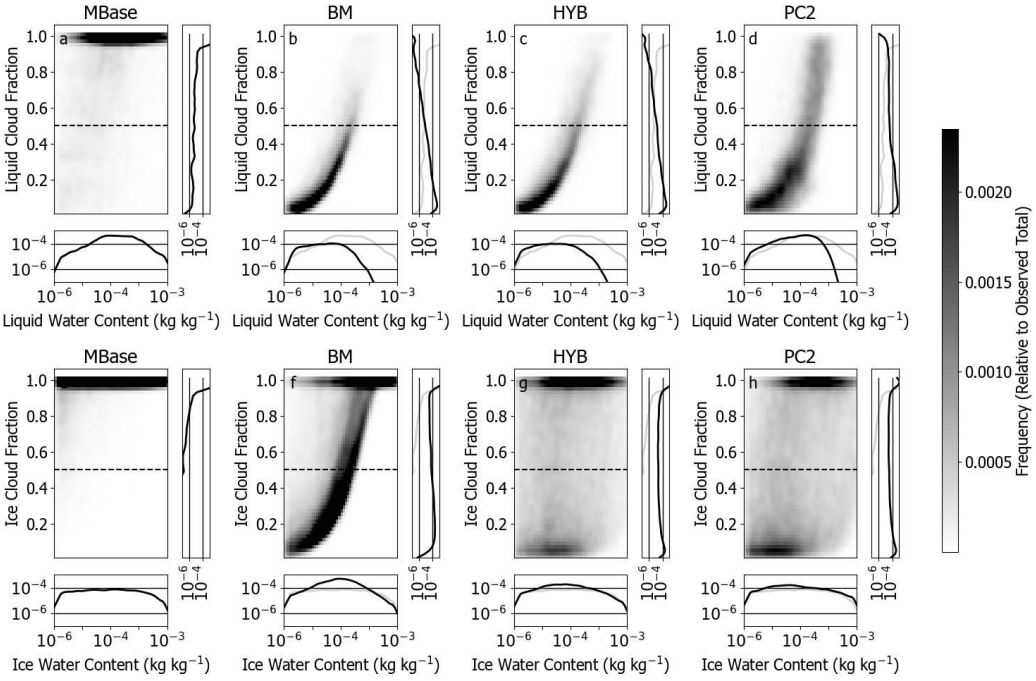

Figure 8. Normalized joint distribution of liquid (cloud + rain) water content versus CF_{liq} (top row) and frozen (ice + snow) water content versus CF_{fro} (bottom row) as obtained from the MBase (a,e) and as simulated by BM (b, f), HYB (c, g) and PC2 (d, h). Shading is on a logarithmic scale. Histograms to the right and the bottom of the joint distributions show the CF and water content distributions, respectively. The grey lines on the simulation histograms denote the observations for reference.

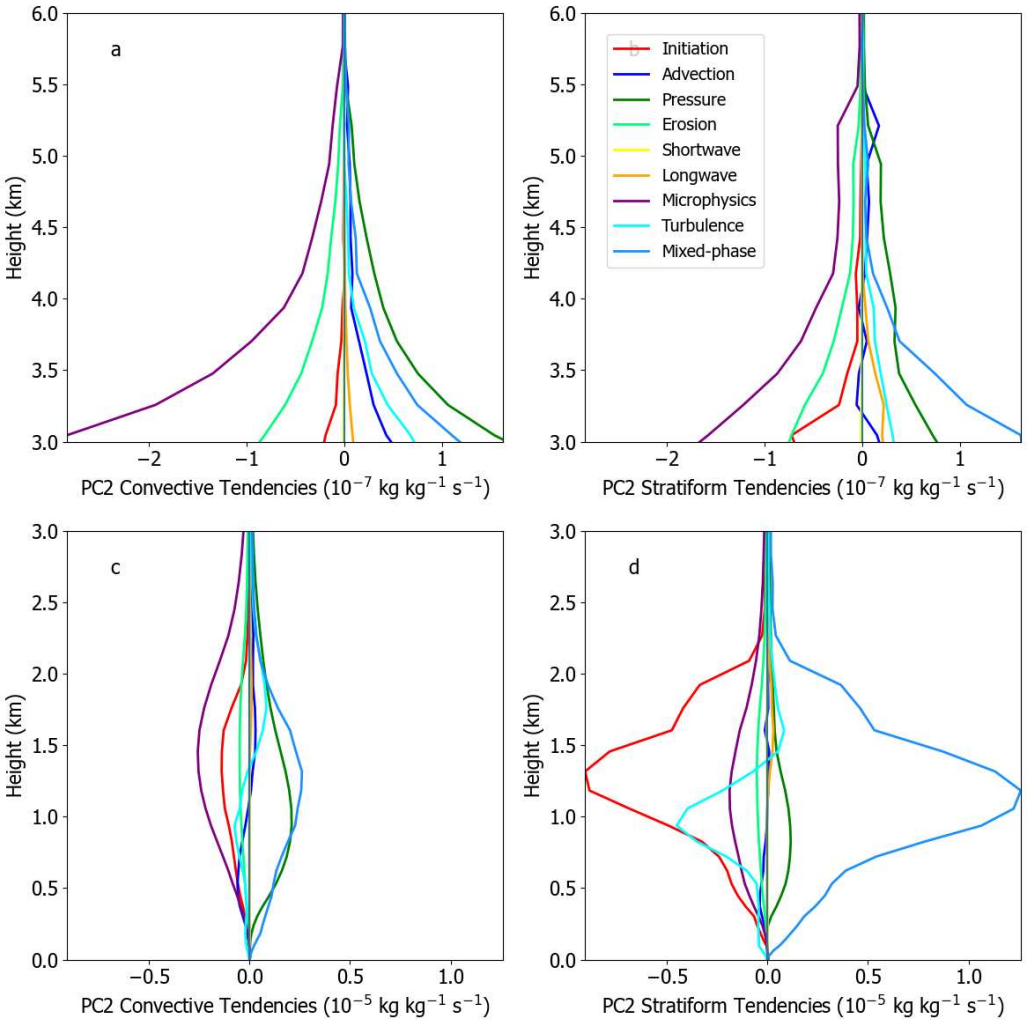

Figure 9. Vertical profiles of the time-averaged liquid water source terms in PC2 at the location of the ANX site for (left) the CNV and (right) the STR CAO events and for mid-level (top) and low-level (bottom) clouds.

A prognostic representation of CF_{fro}, as in HYB (Figure 8g) achieves more spread in the joint q_{fro}-CF_{fro} relation. However, given its diagnostic treatment, the tight relation for liquid as in BM persists. Prognostic treatment of liquid (PC2) increases q_{liq} and CF_{liq}, although even in this case, full cloud cover is seldom reached. We note again that due to the uncertainties in MBase, no firm conclusions should be drawn from the comparison against the observed CF_{liq}. Nevertheless, the enhanced q_{liq}, for which observations are more reliable, is a significant model improvement in PC2 (bottom histograms in Figure 8d).

Given the apparent benefit of prognostic liquid in PC2, it is instructive to investigate the individual liquid water source terms for PC2. These are shown as vertical profiles in Figure 9 . We refer to Wilson et al. (2008) and Furtado et al. (2016) for more detailed descriptions of the source terms. For mid-level clouds (Figure 9a and b), the main sources of q_{liq} are adjustments following the pressure solver and the mixed-phase generation of liquid as described in Furtado et al. (2016). Liquid is mainly consumed by the

microphysics and by cloud erosion, which represents dry air entrainment near cloud edges (Morcrette, 2012). The pressure solver term is particularly associated with updraft regions, while the mixed-phase term becomes dominant in downdraft regions (not shown). Hence, liquid in mid-level updrafts in the CNV mode originate mostly from the pressure term and from upward advection (not shown). At lower tropospheric levels (Figure 9c and d), the STR and CNV mode are more different in terms of PC2 tendencies. The CNV mode has smaller production term magnitudes than the STR mode. Mixed-phase generation of liquid is the dominant source of q_{liq}, in particular in the STR mode. Conversely, PC2 initiation and boundary-layer mixing remove a lot of the liquid produced by the mixed-phase term. Note that the initiation term is a source term if some cloud ought to form in a clear-sky grid box, but becomes a sink when some cloud ought to be removed from an overcast grid box, consistent with the underlying BM PDF assumptions.

It should be mentioned that there are some aspects of the design of PC2 that warrant further research. The PC2 pressure forcing and the mixed-phase generation of liquid terms both act on cloud fields at the end of the pressure solver, while ideally they should be integrated in a single term. In its current formulation, they both act to remove some supersaturation, but with inconsistent underlying assumptions of their SD PDF. More detailed analysis is required to verify whether the improved q_{liq} in PC2 occurs for the right reasons.

Additional sensitivity tests removing the erosion term and the mixed-phase generation of liquid term, respectively led to a big increase and decrease by a factor of 2 of q_{liq} in PC2. Revisiting the formulations of cloud erosion in the context of mixed-phase clouds would be highly recommended to explore potential further improvements in PC2.

4.4 Microphysical Considerations

While previous analysis shows that the MacP matters for mixed-phase cloud-radiative effects, q_{liq} is widely underestimated in all configurations, consistent with previous studies (Klein et al., 2009; Solomon et al., 2009; Forbes & Ahlgrimm, 2014; Field et al., 2017).

From Figure 6, simulations struggle to capture the observed partitioning of frozen condensate in small crystals and large snow aggregates. This is consistent with Solomon et al. (2009), finding underestimated ice number concentrations in mixed-phase simulations using a two-moment MicP.

While no research aircraft were flown during COMBLE to obtain direct measurements of ice number concentrations, some additional indirect evidence of too large precipitation particles can be obtained from the Laser disdrometer (D. Wang et al., 2022), as shown in Figure 10. From this figure, raindrop numbers for a given rain rate tend to be too small in the model (and hence raindrops too large). These large raindrops likely originate from large melting snowflakes in CASIM.

While additional observational evidence for the excessive particle sizes is highly desirable, two additional experiments investigate the contribution of two major snow growth mechanisms to the inception of the particle size bias and their role in the q_{liq} depletion.

Vastly reducing riming growth (HYB-RIM), only alleviates the q_{liq} bias slightly (Figure 2a), while also reducing the q_{fro} considerably (Figure 2c). From Table 6, virtually all ice and snow in this experiment is produced through depositional growth. Despite the reduction in snow production through riming, HYB-RIM has very similar profiles for snow and ice water content and number concentration compared to HYB (Figure 11). The representation of riming therefore does not appear to have a major impact on the ice and snow properties in these simulations, or on the q_{liq} bias.

In contrast, reducing the depositional growth rate of snow (HYB-DEP) brings the LWP bias close the observational uncertainty range in both modes (Figure 2a). Unsur-

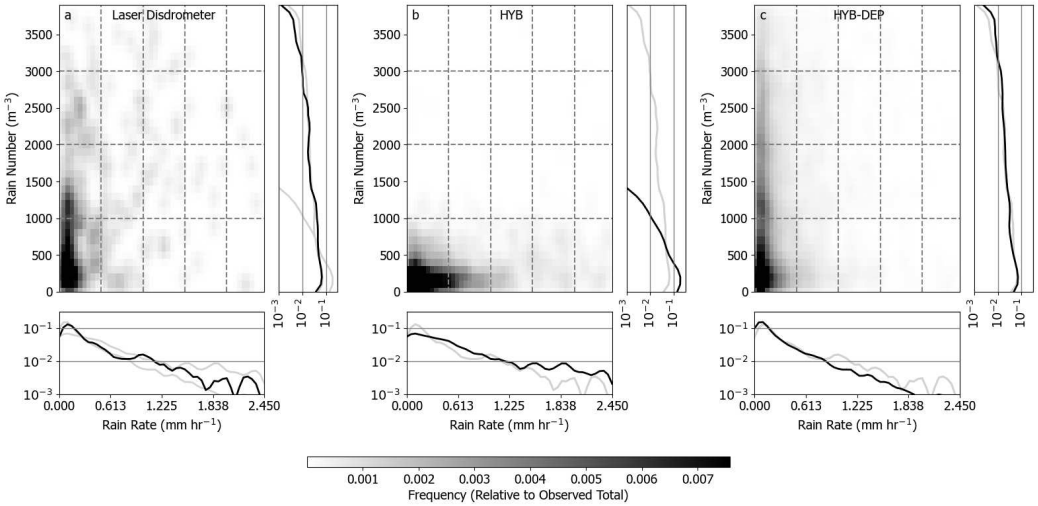

Figure 10. Joint distribution of rain rate and raindrop number concentration as (a) observed during COMBLE with the Laser disdrometer, and as simulated by (b) HYB and (c) HYB-DEP. Note that observations were screened for the occurrence of any frozen or mixed-phase precipitation using the reported World Meteorological Organization weather codes, and the model only contains liquid rain rates. Shading is on a logarithmic scale. Histograms to the right and the bottom of the joint distributions are for the rain number and rain rates respectively. Black lines are for the observations in the left panel, for HYB in the middle panel and for HYB-DEP in the right panel. Grey lines show the other panel's histograms for reference.

prisingly, riming becomes relatively more important in this experiment (Table 6). HYB-DEP experiences greatly increased snow number concentrations, and surprisingly, an increase in snow water content (Figure 11c, f, i, l). The latter can be explained by the more numerous, smaller particles and hence smaller fall speeds. Total snow production in this experiment is indeed much reduced compared to the other experiments (Table 6). The one order-of-magnitude larger snow number concentrations imply much smaller particles in HYB-DEP. Figure 6 suggests that this is in better agreement with observations. From Figure 10c, the smaller snow particles now also melt into smaller (more numerous) raindrops, again in closer agreement to the observations and providing additional evidence for improved precipitation size distributions in HYB-DEP.

Thus, HYB-DEP suggests that depositional growth rates in the default CASIM configurations may be excessive and that this possibly plays a significant role in liquid water depletion. Indeed, if the ice phase grows too excessively through deposition, much of the ice supersaturation will be consumed, depleting liquid water through the Bergeron process. Further research into the role of snow depositional growth in CAO events is needed, e.g. through comparison against observations of snow and ice particle habits and particle fall speeds. Indeed, simulating incorrect particle size distributions, habits or surface area, will greatly affect the depositional growth rates in the simulations.

Note that HYB-DEP alleviates the dry bias in the lower-troposphere somewhat in the simulations shown here (likely due less to vigorous consumption of water vapour through depositional growth). However, a cold, wet bias now emerges in the mid-levels. Further, while the cloud-radiative properties largely improve in the STR mode, excessive ice in the CNV mode in HYB-DEP (Figure 2c), leads to a significant negative SW radiation bias (Figure 4a). HYB-DEP also produces smaller surface precipitation rates than the

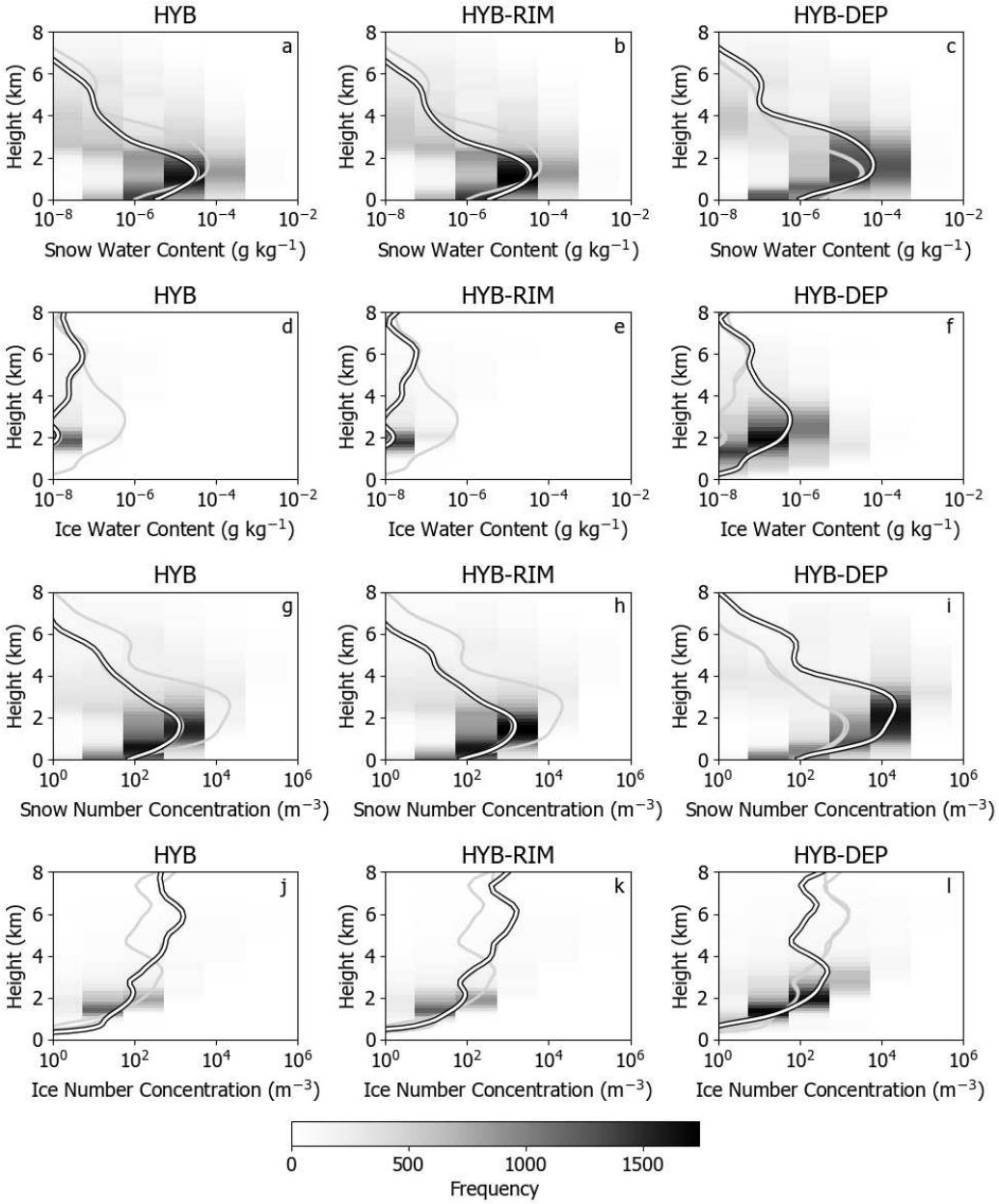

Figure 11. Contoured Frequency by Altitude Diagrams (CFADS) of (a-c) snow and (d-f) ice water content and (g-i) snow and ice (j-l) number concentration, for all CAO events at the location of the ANX site. Shown are (left) HYB, (middle) HYB-RIM and (right) HYB-DEP. Shading of relative frequencies is on a logarithmic scale. The black-and-white line denotes the domain- and time average profile of the experiment indicated in the title of the panel, while the grey lines show the profile for the other two experiments for reference.

other configurations, at the lower boundary of the observational uncertainty (Table 5). It is interesting to note that PC2, which also produces enhanced LWP, has a similar riming fraction (larger than most other configurations) than HYB-DEP, although its total production rates are much larger.

5 Discussion and Conclusions

Cold air outbreak clouds (CAO) in high latitude regions pose a major challenge for regional Numerical Weather Prediction and climate models. These clouds often contain substantial liquid despite their low temperatures. The production and maintenance of this liquid involves processes at scales ranging from microphysics, to meter-scale turbulent motions and mesoscale cloud organization.

Most models underestimate supercooled liquid in these mixed-phase conditions and consequently misrepresent cloud-radiative effects and possibly climate-change feedbacks over polar regions (Klein et al., 2009; Solomon et al., 2009; Field et al., 2017). Many studies so far have looked at aerosol (Tornow et al., 2021), microphysics (Solomon et al., 2009; Abel et al., 2017) and boundary-layer processes (Field, Cotton, et al., 2014; Field et al., 2017) to explore the reasons for this model deficiency. Some of these studies suggest that the maintenance of liquid in mixed-phase clouds involves subgrid-scale phase separation of liquid and ice, justifying a closer look at the role of the subgrid macrophysics parameterization (MacP).

This is done through km-scale simulations of 47 well-observed CAO cases during the Cold-Air Outbreaks in the Marine Boundary Layer Experiment (COMBLE) in Norway in 2020 (Geerts et al., 2022). All simulations were performed using a recently developed two-moment microphysics scheme, CASIM (Field et al., 2022) in the UK Met Office Unified Model. The continuous ground-based observations, collected at one site over a multi-month period, provide a unique opportunity for a more statistically robust analysis, compared to single case studies observed by aircraft.

Key findings of this study are:

- (1) Even at km-scale, using a MacP is beneficial to represent cloud-radiative effects in simulations of convective and stratiform CAO events, although many MacPs suffer from compensating biases. This finding is likely very resolution-dependent and it would be interesting to explore whether this still holds at sub-kilometer scales.
- (2) All diagnostic MacPs using the default CASIM microphysics scheme largely underestimate liquid water path (LWP). In stratiform CAO events, they also underestimate the ice water path (IWP), leading to large SW radiation biases, but in convective CAO events, the IWP is generally better captured, producing smaller radiation errors.
- (3) Of all MacP experiments, only the introduction of prognostic liquid fraction could substantially reduce the LWP bias, but even this configuration produced less than 50% of the observed LWP in convective and Stratiform CAO events. Furthermore, given inconsistent subgrid assumptions for moisture variability between its different components (asymmetric bimodal Gaussian distributions for the initiation term, and symmetric combinations of power law distributions for the other terms), it is hard to determine whether the increase in LWP occurs for the right reasons.
- (4) Phase-overlap and subgrid variability assumptions hardly affect the LWP bias, but produce vastly different IWP and high cloud cover, hence producing very different cloud-radiative effects, in particular in stratiform CAO conditions. Generally, more advanced approaches and minimal overlap produce better cloud-radiative effects and less error compensation.

- (5) Boundary layers in the simulations are too cold, destabilizing the lower troposphere over the relatively warm ocean surface. This could possibly feed back into too rapid glaciation, further exacerbating the cold bias and destabilizing the atmosphere. While it is hard to disentangle cause and effect, this possible liquid-depleting feedback mechanism should be explored in future studies.
- (6) The largest improvement in LWP occurs when greatly reducing the snow depositional growth rate, leading to smaller and more numerous snow particles, in better agreement with observations. Depositional growth in the other configurations appears to consume much of the ice supersaturation, depleting liquid water through the Bergeron process.

This study also highlights the need for a greater effort to unify assumptions across parameterizations. This is an important prerequisite for reducing error compensation and to help disentangle the origin of model biases. For instance, it is hard to understand the reason for enhanced LWP in the prognostic scheme tested here, since various components of the scheme have inconsistent subgrid-variability assumptions.

A new, hybrid cloud scheme is proposed in this paper, building on Van Weverberg and Morcrette (2022). This scheme uses diagnostic liquid cloud, following Van Weverberg, Morcrette, Boutle, Furtado, and Field (2021), but prognostic ice-phase cloud with cloud-fraction tendencies from each microphysical source term. This scheme was found to replicate some of the improved joint cloud fraction - cloud water distributions of the fully prognostic scheme, but at a reduced cost and complexity. However, it was not capable of significantly improving the LWP.

This study accentuates the need for in-situ observations of ice particle sizes and habits. Indeed, more observational evidence would shed light on the role of ice nuclei, snow habit and particles sizes on the possibly excessive depositional growth rate. A caveat of our results is the large observational uncertainty associated with the hydrometeor/phase classification algorithm or other methods to differentiate phase (i.e., simple temperature-based partitioning). A detailed evaluation of these phase identifications against in-situ aircraft observations into regimes similar to those found during COMBLE is highly desirable to enhance the confidence in the results/designations presented here. Nevertheless, it should be emphasized that the evaluation against ground-based measurements during 47 convective and stratiform CAO events during COMBLE has confirmed model biases reported previously for single case studies in a more statistically robust manner.

While this study is based on the UK Unified Model, the results should be relevant to other regional models, given that liquid water depletion in mixed-phase clouds is a widely shared model bias (Klein et al., 2009; Abel et al., 2017; Field et al., 2017). It would be interesting to repeat this analysis for other models and other microphysics schemes to investigate the robustness of the insensitivity to the macrophysics parameterization, and the sensitivity to the formulation of the snow depositional growth rate.

Appendix A Consistent Treatment of Saturation-Departure in Micro- and Macrophysics

The CASIM microphysics and the macrophysics parameterizations (MacPs) were developed independently of one another, and their assumptions are not necessarily consistent. This Appendix discusses the improved coupling between all MacPs and CASIM in all configurations discussed in this paper. Firstly, all MacPs assume liquid ($\mathrm{CF_{liq}}$) and ice clouds ($\mathrm{CF_{fro}}$) to be minimally overlapped, while the cloud-overlap in CASIM is parameterized as follows:

$$CF_{mix} = \min(1, \max(0, \Phi \min(CF_{liq}, CF_{fro}) + \max(0, (1-\Phi)(CF_{liq} + CF_{fro} - 1)))) \quad \text{(A1)}$$

Where the overlap parameter $\Phi = 0$ for a minimal overlap (minimal mixed-phase) and $\Phi = 1$ for a maximal overlap (maximal mixed-phase). The default RAL3 configuration uses a value of $\Phi = 0.5$, implying a random phase overlap. This leads to the following combined cloud fraction CF_{tot}:

$$CF_{tot} = \max(0, \min(1, CF_{liq} + CF_{fro} - CF_{mix})) \tag{A2}$$

All configurations in this manuscript use a consistent phase overlap in all MacPs and CASIM with $\Phi = 0$ (minimal overlap), unless mentioned otherwise.

BM and PC2 initiation calculate the liquid water content (q_{liq}) and liquid water fraction (CF_{liq}) assuming a subgrid SD variability following Field, Hill, et al. (2014) and Furtado et al. (2016). The q_{liq} and CF_{liq} are reduced in the presence of ice, leaving some supersaturation for ice depositional growth in CASIM, as outlined in Field, Hill, et al. (2014); Furtado et al. (2016). This could be seen as a more explicit treatment of the Bergeron process. Smith assumes a triangular subgrid SD distribution, with the width determined by a prescribed profile of critical relative humidity. However, by default CASIM only accounts for grid-mean ice-supersaturation and ignores any subgrid variability assumed by the MacP. This could lead to a situation where a grid box is diagnosed partially liquid-saturated by e.g. BM, but treated as entirely ice-sub-saturated by CASIM. Therefore, a modification was made to pass on the subgrid ice-supersaturation from the MacPs to CASIM, calculated as follows (note that in case of Smith q_{liq} is the entire supersaturation w.r.t liquid):

$$q_{si} = q_{st} - q_{liq} \tag{A3}$$

Where q_{liq} is calculated as in Furtado et al. (2016) for BM and PC2. In BM and PC2, the grid-box total supersaturation q_{st} is given by

$$q_{st} = \begin{cases} 0 & \mu_{ice} < -3\sigma_{turb} \\ \frac{\mu_{ice}}{2}(1 + erf(\frac{\mu_{ice}}{\sqrt{2}\sigma_{turb}})) + \frac{\sigma_{turb}}{\sqrt{2\pi}} \exp \frac{-{\mu_{ice}}^2}{2{\sigma_{turb}}^2} & \\ & -3\sigma_{turb} <= \mu_{ice} <= 3\sigma_{turb} \\ \mu_{ice} & \mu_{ice} > 3\sigma_{turb} \end{cases}$$

This Equation is obtained by integrating under the saturated part of a Gaussian ice-saturation departure PDF with first moment μ_{ice} and second moment σ_{turb} (see Van Weverberg, Morcrette, Boutle, Furtado, and Field (2021) for more detail). The grid-mean total ice saturation departure (SD) μ_{ice} is given by:

$$\mu_{ice} = a_i(\overline{q_T} - q_{si}(\overline{T_{fro}})) \tag{A4}$$

Where q_T represents the total (liquid + frozen) water content, and $q_{si}(\overline{T_{fro}})$ is the saturation specific humidity w.r.t ice using the frozen temperature ($T_{fro} = T - \frac{L_v}{c_p} q_{liq} - \frac{L_f}{c_p} q_{ice}$). The turbulence-based variance ${\sigma_{turb}}^2$ is given by:

$$\sigma_{turb}^2 = (1/2)(a_i \alpha_{ice} \frac{g}{c_p})^2 {\sigma_w}^2 \tau_L \tau_E \tag{A5}$$

α_{ice} is the linear approximation to the local change of q_{si} w.r.t. frozen temperature T_{fro} and the factor a_i accounts for changes in $q_{si}(T_{fro})$ due to latent heating, given by:

$$a_i = (1 + \alpha_{ice}\frac{L_f}{c_p})^{-1} \tag{A6}$$

The time scales for the Lagrangian turbulence de-correlation (τ_L) and homogenization of the turbulent layer (τ_E) are calculated as outlined in Van Weverberg, Morcrette, Boutle, Furtado, and Field (2021). Passing on the subgrid ice supersaturation to CASIM allows depositional growth to take place before a grid-box mean becomes supersaturated and should be more physically consistent with the assumptions made in the MacP.

In Smith, the grid-box total supersaturation q_{st} is given by integrating under a triangular subgrid PDF, yielding (Smith, 1990):

$$q_{st} = \begin{cases} 0 & Q_N <= -1 \\ \frac{b_s}{6}(1+Q_N)^3 & -1 < Q_N <= 0 \\ b_s(Q_N + \frac{1}{6}(1-Q_N)^3) & 0 < Q_N <= 1 \\ b_s Q_N & 1 < Q_N \end{cases}$$

Where Q_N is defined as:

$$Q_N = \frac{a_i(\overline{q_T} - q_{si}(\overline{T_{fro}}))}{b_s} \tag{A7}$$

and the width of the triangular PDF b_s is given by:

$$b_s = a_i(1 - RH_{crit}q_{si}(\overline{T_{fro}})) \tag{A8}$$

It should be said that the impact of the above modifications was generally small with a tendency for simulations to be more glaciated, but little impact on the CF and overall water content. The increased glaciation is not surprising given that depositional growth now can take place before grid-mean saturation is attained.

Appendix B Prognostic cloud fraction tendencies associated with microphysics

Prognostic CF schemes, such as PC2 (Wilson et al., 2008), calculate tendencies of CF from each model component changing the thermodynamic state of the atmosphere. Originally, the PC2 scheme was coupled to the one-moment microphysics scheme of Wilson and Ballard (1999), with CF_{fro} tendencies from nucleation, deposition, melting and sublimation, and liquid cloud (CF_{liq}) tendencies from nucleation, deposition and evaporation. This appendix provides PC2 tendency formulations when coupled to CASIM (Field et al., 2022). Note that while CASIM includes two separate categories for small ice and large aggregates, it only includes a single CF_{fro} for both categories combined.

Homogeneous (pihom) and heterogeneous nucleation (pinuc) are a CF_{liq} sink and a CF_{fro} source, proportional to the q_{liq} fraction that has been converted to ice at the timestep:

$$\Delta CF_{fro} = CF_{liq}(\Delta q_{ice}(pihom) + \Delta q_{ice}(pinuc))/q_{liq} \tag{B1}$$

Melting of ice (pimlt) and snow (psmlt) are a sink for CF_{fro} and are assumed to be added to the rain category (hence not affecting CF_{liq}): Again, the fraction tendency is proportional to the water content tendencies:

$$\Delta CF_{fro} = -CF_{fro}(\Delta q_{rain}(pimlt) + \Delta q_{rain}(psmlt))/(q_{ice} + q_{snow}) \tag{B2}$$

Riming of cloud water onto ice (piacw), snow (psacw) or graupel (pgacw) is a CF_{liq} sink, but not a CF_{fro} source, since riming only occurs in the mixed-phase grid-box portion, which, by definition is already covered in frozen condensate: The CF_{liq} sink is proportional to the riming rate:

$$\Delta CF_{liq} = -CF_{mix}(\Delta q_{ice}(piacw) + \Delta q_{snow}(psacw) + \Delta q_{graupel}(pgacw))/q_{liq} \quad \text{(B3)}$$

Autoconversion (praut) and accretion (pracw) of cloud by rain reduces CF_{liq}, proportional to the q_{liq} loss:

$$\Delta CF_{liq} = -CF_{liq}(\Delta q_{rain}(piacw) + \Delta q_{rain}(pracw))/q_{liq} \quad \text{(B4)}$$

For depositional growth of ice (pidep) and snow (psdep), a slightly different approach was taken, and it is assumed that these processes fill the entire supersaturated portion of the grid box with ice cloud:

$$\Delta CF_{fro} = area_{si}(\Delta q_{ice}(pidep) + \Delta q_{snow}(psdep)/q_{si} \quad \text{(B5)}$$

where q_{si} is calculated as in Equation A3 and the supersaturated grid-box fraction ($area_{si}$) is calculated as:

$$area_{si} = \begin{cases} 0 & \mu_{ice} < -3\sigma_{turb} \\ \frac{1}{2}(1 + erf(\frac{\mu_{ice}}{\sqrt{2}\sigma_{turb}})) & -3 * \sigma_{turb} <= \mu_{ice} <= 3\sigma_{turb} \\ 1 & \mu_{ice} > 3\sigma_{turb} \end{cases}$$

Where the first (μ_{ice}) and second (σ_{turb}) moments are given by Equation A4 and A5.

The CF_{fro} sedimentation tendency is described in Walters et al. (2019), taking into account vertical windshear.

Acknowledgments

Data were obtained from the Atmospheric Radiation Measurement (ARM) User Facility, a U.S. Department of Energy (DOE) Office of Science user facility managed by the Office of Biological and Environmental Research. The work of K. Van Weverberg was supported by the Met Office Weather and Climate Science for Service Partnership (WCSSP) Southeast Asia as part of the Newton Fund, and by the U.K. Civil Aviation Authority.

Appendix C Open Research

This work is conducted in the framework of the UK Unified Model, for which the code is available on a password-protected repository (*Met Office Science Repository*, 2022). All model output fields used in the analysis presented in this paper is available via `https://zenodo.org/record/7422826`. All observational sources (M. Wang et al., 2022; Zhang & Levin, 2022; Jensen et al., 2022; Xiao & Xie, 2022; Kyrouac & Shi, 2022; D. Wang et al., 2022) are available via `https://adc.arm.gov/discovery/`.

References

Abel, S., Boutle, I., Waite, K., Fox, S., Brown, P., Cotton, R., ... Bower, K. (2017). The role of precipitation in controlling the transition from stratocumulus to cmulus clouds in a northern hemisphere cold-air outbreak. *jas*, *74*, 2293-2314. (doi:10.1175/JAS-D-16-0362.1)

Baran, A., Hill, P., Walters, D., Hardiman, S., Furtado, K., Field, P., & Manners, J. (2016). The impact of two coupled cirrus microphysics-radiation parameterizations on the temperature and specific humidity biases in the tropical tropopause layer in a climate model. *Journal of Climate*, *29*, 5299-5316.

Boutle, I., Eyre, J., & Lock, A. (2014). Seamless stratocumulus simulations across the turbulent gray zone. *mwr*, *142*, 1655-1668.

Boutle, I., Finnenkoetter, A., Lock, A., & Wells, H. (2016). The london model: Forecasting fog at 333 m resolution. *qj*, *142*, 360-371.

Bush, M., Allen, T., & coauthors. (2020). The first met office unified model - jules regional atmosphere and land configuration, ral1. *geodev,*, *13*, 1999-2029.

Bush, M., Boutle, I., Edwards, J., Finnenkoetter, A., Franklin, C., Hanley, K., ... Weeks, M. (2022). The second met office unified model/jules regional atmosphere and land configuration, ral2. *Geoscientific Model Development Discussions*, *2022*, 1-35. Retrieved from `https://gmd.copernicus.org/preprints/gmd-2022-209/` doi: 10.5194/gmd-2022-209

Cadeddu, M. P., Ghate, V. P., & Mech, M. (2020). Ground-based observations of cloud and drizzle liquid water path in stratocumulus clouds. *Atmospheric Measurement Techniques*, *13*(3), 1485-1499. doi: 10.5194/amt-13-1485-2020

Cahalan, R., Ridgway, W., Wiscombe, W., & Bell, T. (1994). The albedo of stratocumulus clouds. *jas*, *51*, 2434-2455.

Dipankar, A., Webster, S., Sun, X., Sanchez, C., North, R., Furtado, K., ... Barker, D. (2020). Singv: A convective-scale weather forecast model for singapore. *qj*, *146*, 4131-4146.

Edwards, J., & Slingo, A. (1996). Studies with a flexible new radiation code. i: Choosing a configuration for a large-scale model. *qj*, *122*, 689-719.

Eloranta, E. (2005). *High spectral resolution lidar, in lidar; range-resolved optical remote sensing of the atmosphere*. New York: Springer-Verlag.

Field, P., Brozkova, R., Chen, M., Dudhia, J., Lac, C., Hara, T., ... McTaggert-Cowan, R. (2017). Exploring the convective grey zone with reginoal simulations of a cold air outbreak. *qj*, *143*, 2537-2555. (doi:10.1002/qj.3105)

Field, P., Cotton, R., McBeath, K., Lock, A., Webster, S., & Allan, R. (2014). Improving a convection-permitting model simulation of a cold air outbreak. *qj*, *140*, 124-138. (doi:10.1002/qj.2116)

Field, P., Hill, A., Furtado, K., & Korolev, A. (2014). Mixed-phase clouds in a turbulent environment. part ii: Analytic treatment. *qj*, *140*, 870-880.

Field, P., Hill, A., Furtado, K., Wilkinson, J., Miltenberger, A., Gordon, H., ... Van Weverberg, K. (2022). Implementation and testing of a double moment cloud microphysics in uk met office regional numerical weather prediction. *mwr*, *accepted*, 0-0.

Forbes, R., & Ahlgrimm, M. (2014). On the representation of high-latitude bound-

ary layer mixed-phase clouds in the ecmwf global model. *mwr*, *142*, 3425-3445. (doi:10.1175/MWR-D-13-00325.1)

Frey, W., & Kay, J. (2018). The influence of extratropical cloud phase and amount feedbacks on climate sensitivity. *Climate Dynamics*, *50*, 3097-3116.

Furtado, K., & Field, P. (2017). The role of ice microphysics parametrizations in determining the prevalence of supercooled liquid water in high-resolution simulations of a southern ocean midlatitude cyclone. *jas*, *74*, 2001-2021.

Furtado, K., Field, P., Boutle, I., Morcrette, C., & Wilkinson, J. (2016). A physically based subgrid parameterization for the production and maintenance of mixes-phase clouds in a general circulation model. *jas*, *73*, 279-291.

Geerts, B., Giangrande, S., & coauthors. (2022). The comble campaign: A study of marine boundary layer clouds in arctic cold-air outbreaks. *bams*, *103*, 1371-1389. (doi:10.1175/BAMS-D-21-0044.1)

Good, S., Fiedler, E., Mao, C., Martin, M., Maycock, A., Reid, R., ... Worsfold, M. (2020). The current configuration of the ostia system for operational production of foundation sea surface temperature and ice concentration analyses. *Remote Sensing*, *12*, 720. (doi:10.3390/rs1204720)

Gruetzun, V., Quaas, J., Morcrette, C., & Ament, F. (2013). Evaluating statistical cloud schemes: What can we gain from ground-based remote sensing? *jgr*, *118*, 10507-10517.

Hughes, K., Ross, A., Vosper, S., Lock, A., & Jemett-Smith, B. (2015). Assessment of valley cold pools and clouds in a very high resolution numerical weather prediction model. *geodev,*, *8*, 3105-3117.

Illingworth, A., Hogan, R., & coauthors. (2007). Cloudnet - continuous evaluation of cloud properties in seven operational models using ground-based observations. *bams*, *88*, 883-898.

Jensen, M., Giangrande, S., Fairless, T., & Zhou, A. (2022). *Interpolated sonde (interpolatedsonde) [dataset].* doi: 10.5439/1095316

Jia, Y., Li, Z., Xu, C., Jin, S., & Deng, H. (2020). A comparison of precipitation measurements with a pws100 laser sensor and a geonor t-200b precipitaion gauge at a nival glacial zone in eastern tianshan, central asia. *Atmosphere*, *11*, 1079.

Kalesse, H., de Boer, G., Solomon, A., Oue, M., Ahlgrimm, M., Zhang, D., ... Protat, A. (2016). Understanding rapid changes in phase partitioning between cloud liquid and ice in stratiform mixed-phase clouds: An arctic case study. *Monthly Weather Review*, *144*(12), 4805 - 4826. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/144/12/mwr-d-16-0155.1.xml` doi: https://doi.org/10.1175/MWR-D-16-0155.1

Klein, S., & Hartmann, D. (1993). The seasonal cycle of low stratiform clouds. *Journal of Climate*, *6*, 1588-1606. (doi:10.1175/1520-0442(1993)006)

Klein, S., McCoy, R., & coauthors. (2009). Intercomparison of model simulations of mixed-phase clouds observed during the arm mixed-phase arctic cloud experiment. i: Single-layer cloud. *qj*, *125*, 979-1002. (doi:10.1002/qj.416)

Korolev, A., & Field, P. (2008). The effect of dynamics on mixed-phase clouds: Theoretical considerations. *jas*, *65*, 66-86.

Korolev, A., McFarquhar, G., Field, P. R., Franklin, C., Lawson, P., Wang, Z., ... et al. (2017). Mixed-phase clouds: Progress and challenges. *Meteorological Monographs*, *58*. doi: 10.1175/amsmonographs-d-17-0001.1

Kyrouac, J., & Shi, Y. (2022). *Surface meteorological instrumentation (met) [dataset].* doi: 10.5439/1786358

Manners, J., Edwards, J., Hill, P., & Thelen, J. (2018). *Socrates (suite of community radiative transfer codes based on edwards and slingo)* (Tech. Rep.). Met Office, UK. (Available online at: https://code.metoffice.gov.uk/trac/socrates)

McFarquhar, G., & coauthors. (2011). Indirect and semi-direct aerosol campaign: The impact of arctic aerosls on clouds. *bams*, *92*, 183-201.

Met office science repository. (2022). (`https://code.metoffice.gov.uk/trac/home`)

Miller, M., Nitschke, K., Ackerman, T., Ferrell, W., Hickmon, N., & Ivey, M. (2016). The arm mobile facilities. the atmospheric radiation measurement (arm) program: The first 20 years. *Meteor. Monogr.*, *57*.

Morcrette, C. J. (2012). Improvements to a prognostic cloud scheme through changes to its cloud erosion parametrization. *Atmos. Sci. Let.*, *13*, 95-102. (DOI: 10.1002/asl.374)

Morcrette, C. J., O'Connor, E., & Petch, J. C. (2012). Evaluation of two cloud parametrization schemes using arm and cloud-net observations. *Q. J. Roy. Meteor. Soc.*, *138*, 961-979.

Murray, B., Carslaw, K., & Field, P. (2021). Opinion: Cloud-phase climate feedback and the importance of ice-nucleating particles. *acp*, *21*, 665-679. (doi:10.5194/acp-21-665-2021)

Persson, P., Shupe, M., & Perovich, D. e. a. (2017). Linking atmospheric synoptic transport, cloud phase, surface energy fluxes, and sea-ice growth: observations of midwinter sheba conditions. *Clim Dyn*, *49*, 1341-1364. doi: https://doi.org/10.1007/s00382-016-3383-1

Rauber, R., & Tokay, A. (1991). An explanation for the existence of supercooled water at the top of cold clouds. *jas*, *48*, 1005-1023.

Romatschke, U., & Vivekanandan, J. (2022). Cloud and precipitation particle identification using cloud radar and lidar measurements: Retrieval technique and validation. *Earth and Space Science*, *9*(5), e2022EA002299. doi: https://doi.org/10.1029/2022EA002299

Romps, D. M. (2017, December). Exact expression for the lifting condensation level. *Journal of the Atmospheric Sciences*, *74*(12), 3891–3900.

Shi, Y., & Long, C. (2002). *Best estimate radiation flux value added product: Algorithm operational details and explanations.* (Tech. Rep.). US Department of Energy. (Available online at: http://www.arm.gov/publications/tech reports/doe-sc-arm-tr-008.pdf)

Shupe, M. (2007). A ground-based multisensor cloud phase classifier. *Geophysical Research Letters*, *34*(22). doi: 10.1029/2007gl031008

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

Shupe, M., Kollias, P., Persson, O., & McFarquhar, G. (2008). Vertical motions in arctic mixed-phase stratiform clouds. *jas*, *65*, 1304-1322.

Shupe, M., Matrosov, S., & Uttal, T. (2006). Arctic mixed-phase cloud properties derived from surface-based sensors at sheba. *jas*, *63*, 697-711.

Shupe, M. D. (2011). Clouds at arctic atmospheric observatories. part ii: Thermodynamic phase characteristics. *Journal of Applied Meteorology and Climatology*, *50*(3), 645 - 661. doi: https://doi.org/10.1175/2010JAMC2468.1

Silber, I., Verlinde, J., Eloranta, E. W., & Cadeddu, M. (2018). Antarctic cloud macrophysical, thermodynamic phase, and atmospheric inversion coupling properties at mcmurdo station: I. principal data processing and climatology. *Journal of Geophysical Research: Atmospheres*, *123*(11), 6099-6121. doi: https://doi.org/10.1029/2018JD028279

Smith, R. (1990). A scheme for predicting layer clouds and their water content in a general circulation model. *qj*, *116*, 435-460.

Solomon, A., Morrison, H., Persson, O., Shupe, M., & Bao, J. (2009). Investigation of microphysical parameterizations of snow and ice in arctic clouds during m-pace through model-observation comparisons. *mwr*, *137*, 3110-3128. (doi:10.1175/2009MWR2688.1)

Solomon, A., Shupe, M. D., Persson, P. O. G., & Morrison, H. (2011). Moisture and dynamical interactions maintaining decoupled arctic mixed-phase stratocu-

mulus in the presence of a humidity inversion. *Atmospheric Chemistry and Physics*, *11*(19), 10127–10148. doi: 10.5194/acp-11-10127-2011

Stoffel, T. (2005). *Solar infrared radiation station (sirs) handbook.* (Tech. Rep.). US Department of Energy. (Available online at: https://www.arm.gov/publications/tech reports/handbooks/sirs handbook.pdf?id=68)

Tornow, F., Ackerman, A., & Fridlind, A. (2021). Preconditioning of overcast-to-broken cloud transitions by riming in marine cold air outbreaks. *acp*, *21*, 21049-12067.

Turner, D. D., Clough, S. A., Liljegren, J. C., Clothiaux, E. E., Cady-Pereira, K. E., & Gaustad, K. L. (2007). Retrieving liquid wat0er path and precipitable water vapor from the atmospheric radiation measurement (arm) microwave radiometers. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(11), 3680-3690. doi: 10.1109/TGRS.2007.903703

Van Weverberg, K., & Morcrette, C. (2022). Sensitivity of cloud-radiative effects to cloud fractin parametrizations in tropical, midlatitude, and arctic kilometre-scale simulations. *qj*, *148*, 2563-2586.

Van Weverberg, K., Morcrette, C., & Boutle, I. (2021). A bimodal diagnostic cloud fraction parameterization. part ii: Evaluation and resolution sensitivity. *mwr*, *149*, 859-878. (doi:10.1175/MWR-D-20-0230.1)

Van Weverberg, K., Morcrette, C., Boutle, I., Furtado, K., & Field, P. (2021). A bimodal diagnostic cloud fraction parameterization. part i: Motivating analysis and scheme description. *mwr*, *149*, 841-857.

Van Weverberg, K., Morcrette, C., Ma, H., Klein, S., & Petch, J. (2015). Using regime analysis to identify the contribution of clouds to surface temperature errors in weather and climate models. *qj*, *141*, 3190-3206.

Walters, D., Baran, A., & coauthors. (2019). The met office unified model global atmosphere 7.0/7.1 and jules global land 7.0 configurations. *geodev,*, *12*, 1909-1963.

Wang, D., Bartholomew, M., & Shi, Y. (2022). *Laser disdrometer (ld) [dataset].* doi: 10.5439/1779709

Wang, M., Giangrande, S., Johnson, K., & Jensen, M. (2022). *Improved microbase product with uncertainties (microbasekaplus) [dataset].* doi: 10.5439/1768890

Wilson, D., & Ballard, S. (1999). A microphysically based precipitation scheme for the meteorological office unified model. *qj*, *125*, 1607-1636.

Wilson, D., Bushell, A., Kerr-Munslow, A., Price, J., & Morcrette, C. (2008). Pc2: A prognostic cloud fraction and condensation scheme. i: Scheme description. *qj*, *134*, 2093-2107.

Wood, R., & Bertherton, C. (2006). On the relationship between stratiform low cloud cover and lower-tropospheric stability. *Journal of Climate*, *19*, 6425-6432. (doi:10.1175/JCLI3988.1)

Xiao, C., & Xie, S. (2022). *Arm best estimate data products (armbeatm) [dataset].* doi: 10.5439/1333748

Zhang, D., & Levin, M. (2022). *Thermodynamic cloud phase (thermocldphase) [dataset].* doi: 10.5439/1871014

Zhang, D., Vogelmann, A., Kollias, P., Luke, E., Yang, F., Lubin, D., & Wang, Z. (2019). Comparison of antarctic and arctic single-layer stratiform mixed-phase cloud properties using ground-based remote sensing measurements. *jgr*, *124*, 10186-10204. (doi:10.1029/2019JD030673)

Zhao, C., Xie, S., Chen, X., Jensen, M.-P., & Dunn, M. (2014). Quantifying uncertainties of cloud microphysical property retrievals with a perturbation method. *jgr*, *119*, 5375-5385. (doi:10.1002/2013JD021112)