Reviewer Comments, Responses, and Changes to Manuscript (Pages 1-15)
(Revised manuscript with track changes starts on Page 16)

5   *Please note that any references to page or line numbers in our "changes to manuscript" responses refer to the final revised version of the manuscript (without track changes) and reference to any figure numbers in our "changes to manuscript" responses refer to revised figure numbers. Figures included for responses below are referenced as Fig. R1-R5. Figures added to supplemental material are referenced as Fig. S1-S4.*

10 ## Reviewer 1

Specific Comments

1. Comment: The manuscript compared observed TWC and MMD stratified by w and T/height. Ostensibly
15 missing is the comparisons of the total number concentration. The total number concentration should be the most accurate observation and is predicted in all 3 microphysical schemes. Comparison of N_total stratified with w and T and/or TWC will provide more insights in model errors. In a generalized microphysical framework, one moment schemes solve mass equation (first order of mass); two-moment schemes in general solve both mass and number concentration (zeroth and first order of mass). There are also three-moment schemes which solve the zeroth, first
20 and another higher order variable, usually the second moment of mass. Comparisons of the simulated total numbers with observations is essential for all microphysical scheme validations. It can also be carried out in high confidence especially when using in-situ data. I suggest that the authors add N_total with respect to both w and TWC in Fig. 6. And add corresponding N_total in model simulation plots.

25 **Response:** *We agree that total number concentrations are certainly important for microphysics scheme evaluation, however they were not analyzed in detail for two reasons. First, comparisons are really only useful when liquid and ice are separated because liquid dominates the total number concentration when present (e.g., a typical cloud droplet concentration is 100,000 L^{-1} whereas a high ice concentration is 100 L^{-1}). The issue is that liquid and ice number concentrations are not separated in measurements. Additionally, liquid in the form of small*
30 *cloud droplets are common in mixed phase updrafts of the bulk schemes (see Figure 14), and will dominate the number concentration in these regions. Apart from this simulated liquid likely often being erroneous (e.g., from not maintaining liquid supersaturations in the bulk schemes), the cloud droplet number concentration in the bulk schemes is held constant and there are no constraints on the CCN PSD in FSBM other than that this event occurs in a clean, tropical maritime air mass. Therefore, liquid needs to be removed from any number concentration*
35 *comparisons, which is easy in simulations, but difficult in measurements.*

Second, the measured total number concentrations have significant uncertainty. Figure R1, provided by coauthor Alfons Schwarzenboeck, shows differences in three 1-minute composite PSDs (resulting from 5-second samples) using derivation techniques from 3 different institutions. Consistent differences in number concentrations can reach
40 *an order of magnitude for sizes below ~150 µm (note the logarithmic ordinate scaling). These small particles control the total number concentration, but quantifying the observational number concentration uncertainty is a*

nontrivial task, and doing so is beyond the scope of this study. The large differences in PSDs shown in Figure R1 partly result from different techniques used to filter out shattered particles. TWC and MMD measurement have greater certainty than number concentration since the IKP2 measures TWC plus water vapor, where the subtraction of water vapor is the primary source of uncertainty, while MMD is not strongly impacted by differences in particle numbers at diameters smaller than 150 μm. Essentially, mass as a higher moment of the PSD than number concentration is much less impacted by small particles that are the basis for uncertainty. For example, Figure 7 of Leroy et al. (2017) shows that the mean 15% MD as a function of TWC for any given flight is never smaller than 100 μm.

Changes to manuscript: *We have added language on P5 L16-19 in the manuscript to state why total number concentrations are neglected.*
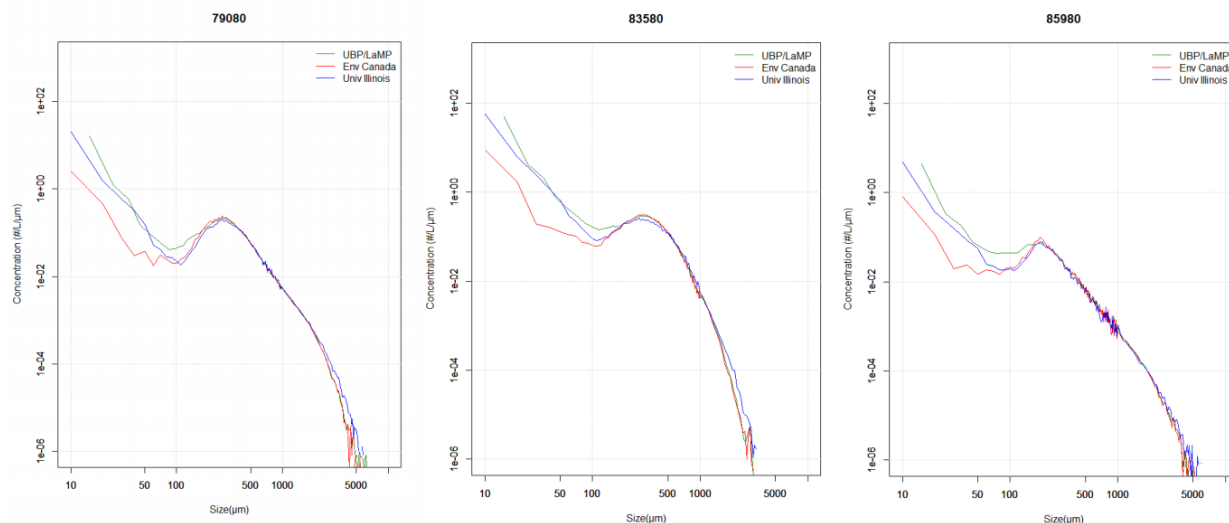

Figure R1. PSDs from 3 different 1-minute composites (resulting from 5-second samples) showing PSDs derived from 3 different institutions. Number concentration is on the ordinate and diameter is on the abscissa. Both axes are logarithmic and each ordinate tick mark indicates 2 orders of magnitude. Green lines are from UBP/LaMP, red lines are from Environment Canada, and blue lines are from the University of Illinois.

2. Comment: Fig. 7, 8 and 9 can be combined to a 3x3 panel figure for easier comparison and discussion.

Response: *We agree that combining these 3 figures into one figure facilitates easier discussion.*

Changes to Manuscript: *Former Figures 7, 8, and 9 have been combined into a 3x3 panel (current Figure 7)*

3. Comment: The color scheme in Fig. 7-9 is also confusing. The convention is that blue has smaller value than red. I was initially confused by the fact that red represents negative and blue positive. I'd suggest the authors to flip the color scheme in these figures.

Response: *We agree that the convention is to use red for positive values and blue for negative values.*

Changes to Manuscript: *All relative difference plots now use red for positive values and blue for negative values (see current Figures 7, 8, S2, S3, and S4).*

4. Comment: What is the rationale of using only data with w>1m/s? Can data with w< 1 m/s be used for model comparisons, too? Will it lead to the same conclusion?

Response: *The original reasoning for using only updrafts was that high IWC regions targeted during the HAIC-HIWC field campaign are typically found in updrafts or regions connected to nearby convective cores, and are hypothesized to form within updrafts and regions detrained out of convective updrafts (see Lawson et al., 1998). However, we agree that this study would be more conclusive by including downdrafts and quiescent regions. This also provides a more statistically robust observational sample size for many significant TWC values (total sample size has increased approximately 8-fold, see Figure 6). An additional constraint was added (see response to Minor Point #2 by Reviewer 2) by which only simulated and observed points with TWC > 0.1 g m^{-3} are included. This value is similar to the expected uncertainty in measured TWC at -40°C (see discussion on P4 L28-33).*

Changes to Manuscript: *The dataset now includes all vertical velocity values with only the following constraints: (1) -60 °C < T < 0 °C and (2) TWC > 0.1 g m^{-3}.*

5. Comment: The biggest problem with the observation is its sampling bias. Due to safety concerns, the airplane must avoid lightning and area with radar reflectivity above 40 dBZ. The authors mentioned this error, but didn't do anything to quantify it. Another problem is that all data from different events are used indiscriminately to compare with a single case simulation. The author should add some analyses to address these uncertainties. For example, Fig. 1 seems to show that on Feb. 18, the airplane sampling might be on the weaker part of the system. This case also seems to have the warmest Tb among the four cases shown. Is this true? Does samplings with high w mainly come from other cases? One way of determine the bias is by plotting sampling sizes for Feb. 18 case only, and compare the result with the same plot using all samples, on a T-w and/or T-TWC plot. Another possibility to address the sampling bias is to combine the 3D C-Pol data with Feb. 18 sampling. For example, one could plot C-POL sampling sizes in T-dBZ space. Then use observed PSD to calculate dBZ for in-situ measurements, and plot sample sizes on a T-dBZ diagram, too. This could roughly show how much/what type of biases existed in airplane sampling.

Response: *There are several difficulties in quantifying the observational sampling bias. Although Fig. 1 shows that the aircraft sampled warmer brightness temperatures in the Feb. 18 system compared to others, this was necessary to observe convective cores within range of the C-POL radar, cores that still had cloud tops reaching up to 14-km altitude as observed by the onboard RASTA W-band radar. It is possible to compare C-POL reflectivity within the flight path to all C-POL reflectivities over the limited range of temperatures observed during this flight, but the most intense convection during this case was before the flight. During the flight, no cells were avoided within the region of interest (the C-POL domain). Therefore, the flight during this event is not representative of*

3

avoidance of high reflectivity regions or lightning, and trustworthy Rayleigh reflectivity was not observed for other cases. Additionally, using full 3-D C-POL reflectivity assumes that the aircraft would observe all regions equally, but this is not true, since the aircraft was specifically targeting regions thought to possibly have high IWC subjectively based on both pilot's radar data and satellite data. Therefore, there is no objective way to quantify
5   *this sampling bias in terms of reflectivity or any other variable. The impact of this bias is limited to the extent possible by controlling for TWC and vertical velocity, and is also the reason that the minimum 90% MD is examined, since it is not impacted by this sampling bias.*

Additionally, Flight 23 on Feb. 18 is not representative of other flights. Attached is a PDF (Figure R2) showing
10   *the contribution of each flight to the dataset used for this study. The 4 simulated events are indicated. On the right ordinate, Figure R2 shows the mean and standard deviation of temperature. Note that although a single flight may have sampled multiple temperature levels, each discrete flight leg flew at an approximately constant temperature level. This PDF signifies the need to use observations from the entire field campaign in order to properly stratify variables by temperature like the data shown in Figure 6. The Feb. 18 case (Flight 23) contains only ~7-8% of the*
15   *total samples, the majority of which remain between -10 and -20 °C. This PDF also show that using data from only the Feb. 18 event would not yield information on relatively colder temperatures, where most observations exist.*

Figures S2-S4 were added in the supplemental material to address the concern of comparing one simulated event with observations from all field campaign events. These figures are similar to Figure 7, but show relative
20   *differences in observed and simulated TWC and MMD as a function of w-T and/or TWC-T bins for 3 additional events (Jan. 23, Feb. 2-3, and Feb. 7 events in Figures S2-S4, respectively) simulated with the bulk schemes. Note that FSBM was only run for the Feb. 18 simulation due to its high computational expense. These supplemental figures show that when controlling for temperature, TWC, and/or w, every simulated event yields very similar differences with observations in MMD-w-T-TWC space (where observations are taken from all flights), particularly*
25   *in the case of MMDs. While there is certainly variability in the peak vertical velocities of simulated events, we control for vertical velocity, and thus this variability does not impact our results.*

We note that the intention of focusing on the Feb. 18 event simulations is that radar (C-POL) observations were available for this event alone, and it contains most of the high TWC and high w observations at relatively warm
30   *temperatures. Figures S2-S4 clearly show that simulating a single event with different microphysics schemes yields much larger differences than simulating different events with the same microphysics scheme, at least in the phase space considered in this study.*

Changes to Manuscript: *Figure R2 has been added to the supplemental material as Figure S1 for extra*
35   *justification of our comparison methodology. Figure S1 is discussed on P4 L7-8 and P7 L19-22. Figures S2-S4 have been added to the supplemental material to show that using a single simulated event is sufficient for comparison with all Darwin observations in the MMD-w-T-TWC phase space considered in this study. A short discussion of Figures S2-S4 is now provided on P11 L32 – P12 L3.*
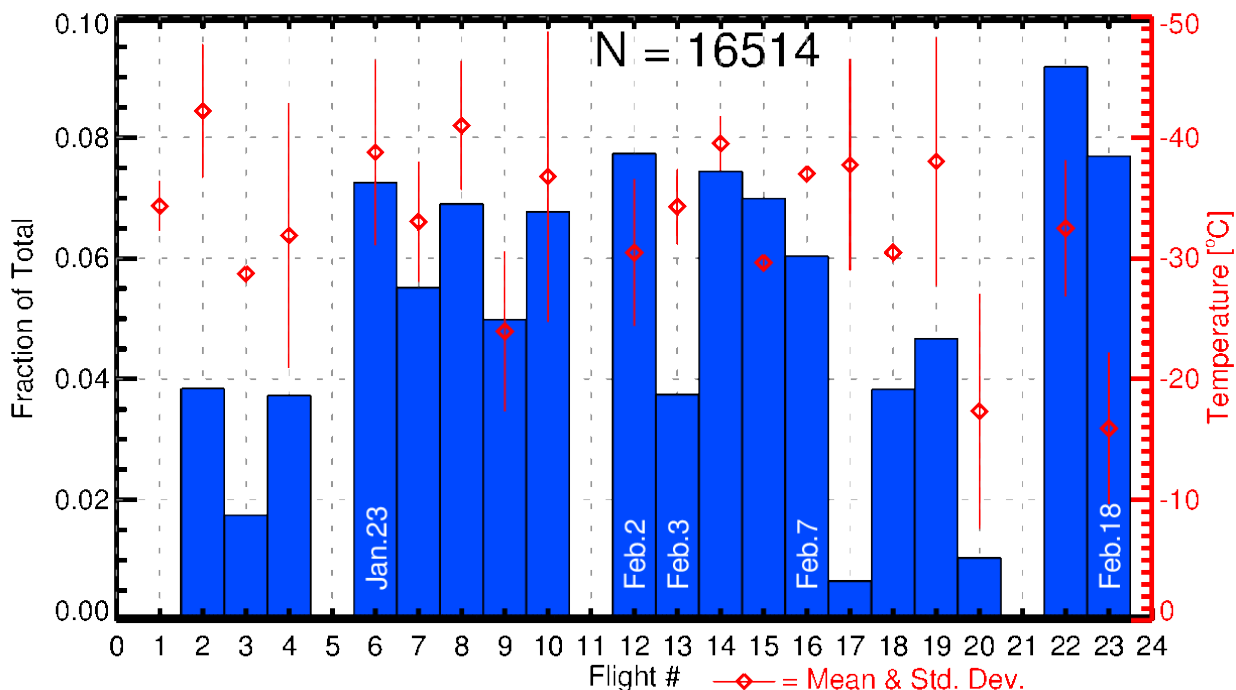
Figure R2. Distribution of flight samples for the Darwin HAIC-HIWC campaign by flight number (blue bars). Red diamonds show the mean temperature for each flight, with lines that show +/- one standard deviation.

6. Comment: The comparisons are made for model output between 18Z on the 18th to 00Z on the 19th, because
5 this time period was considered to represents mature and dissipating stage of the MCS, according to the manuscript. However, Fig. 5 compares reflectivity at 16Z, which is outside the window for PSD comparisons. I suggest the authors to plot C-Pol radar reflectivity CFADs (Yuter, 1995, Mon. Wea. Rev., Vol123, P1941-1963) for the same 6-hour period. Then compare it with CFADs of model simulations.

10 **Response:** *Figure 5 has been revised such that it now shows only the 99th percentile to account for strictly convective radar reflectivities since there is a significant amount of stratiform precipitation across the domain (see Figures 3 and 4). Figure 5a shows the 99th percentile for a time period between 12Z and 18Z on the 18th and Figure 5c shows the 99th percentile profile between 18Z on the 18th and 00Z on the 19th. Figures 5b and 5d show sample sizes normalized by domain area (C-POL domain is smaller than the model domain) for the 12Z-18Z and 18Z-00Z*
15 *time periods, respectively. These two time periods are shown because the most intense convection observed by C-POL was between 12Z-18Z, but the flight took place between 18Z-00Z. Although the most intense convection was not in the C-POL domain during the 18Z-00Z analysis period, it is clear that all simulations produce significantly higher reflectivities aloft compared to C-POL, no matter the time period selected for comparison. 99th percentile profiles and domain-normalized sample sizes are still used as opposed to CFADs. Using percentile profiles on one*
20 *figure allows quantification of differences between observations and simulations that is not possible in CFADs. Moreover, the high IWC regions targeted by the campaign aircraft are generally found in and around convective cores, and regions of modest stratiform reflectivity that would dominate CFADs were not of interest. The large simulated reflectivity biases aloft that are being explored in this study are commonly associated with convective*

regions. This is the bias being investigated with the in situ microphysical and kinematic datasets, and thus comparison of high percentile profiles (e.g. 99th) is more appropriate than CFADs.

Changes to Manuscript: *Importantly, we note that we have used an updated C-POL dataset provided by Alain Protat who has been added as a co-author. We have also gridded these C-POL radial sweep files ourselves to match the simulation horizontal grid spacing (1 km) and found a ~2.5 dB increase, which is why the cross-section figures (Figures 3 and 4) change from the original version, however this increase is not central to any of our conclusions.*

Figure 5 has been revised as discussed in the response above. Discussion of Figure 5 has been revised on P10 L4-15.

Technical Corrections

1. Comment: Please put correct Fig. 4 and Fig. 9.

Changes to manuscript: *The correct version of Figure 4 is now in place. The correct version of former Figure 9 is now represented in the bottom panel of Figure 7 (g-i), as recommended in Specific Comment #2.*

2. Comment: P1, L24: ". . .., differences with observations for a given particle size vary greatly between schemes." I don't understand this statement. Please rephrase.

Changes to manuscript: *Clarified this statement on P1 L22-24.*

3. Comment: P4, L12: "Uncertainty in w calculations is estimated at ~1 ms-1". Please give reference.

Changes to manuscript: *We added a reference to Jorgensen and LeMone (1989) on P4 L13.*

4. Comment: P5, first paragraph: Can the authors list the values of alpha and beta derived from the observations. They can be used to compare with model parameters, and are essential for reproducing the results (e.g., deriving MMD) shown in this study.

Response: *Observed values of α and β are not constant in the observational dataset, as explained on P5 L4-6 and in much more detail in Leroy et al. (2016). For this reason, the > 16,000 m-D coefficient pairs used in the current study are not listed. The β parameter changes during flight since it is related to the exponents of area-size and perimeter-size power laws that are derived directly from OAP images for every 5-second flight sample. The α parameter varies during flight as it is constrained by independent TWC retrievals. Additionally, simulated α and β parameters that are comparable to observations are not possible since observations include all hydrometeor types in derivation of the power law, while simulations have multiple hydrometeor species, each with different α and β parameters, and combining these species to produce the mass size distribution means that the simulated mass-size relationship no longer follows a power law.*

6

Changes to manuscript: *Due to possible confusion, the sentence on P16 L20-23 has been modified to indicate that each 5-second flight sample uses a single m-D power law (not the entire dataset using the same power law/coefficients), and that the assumption that the distribution of mass with diameter is well represented by a power law may not be a good assumption in some situations. However, analyzing composite distributions may alleviate this issue to some degree.*

5. Comment: P5, L5: Deq is defined here as "area equivalent diameter", but later in eqn. 5 is used as "melted equivalent diameter". This is confusing.

Changes to manuscript: *All simulated D_{eq} values are now referenced as $D_{eq,melt}$ in the computation of reflectivity where equivalent melted diameter is used, whereas D_{eq} is now explicitly used to refer to the 2D area equivalent ice diameter, defined as the diameter of a circle with the same area as particle images from the OAPs.*

6. Comment: P5, L27: ACCESS-R is used as the initial and boundary condition, not as "largescale forcing", according to the conventional term usage.

Changes to manuscript: *Changed "large-scale forcing" to "initial and boundary conditions" on P6 L2-3.*

7. Comment: P7 L2, where is the citation for CCN concentration? How high is the boundary layer?

Response: *Perhaps there was some confusion in the values listed that give a vertical representation of the initial aerosol concentration profile in the FSBM scheme. These values are the default values used by FSBM to represent a maritime (pristine) air mass (set using FCCNR_MAR in module_mp_fast_sbm.F, the WRF microphysics module for the FSBM scheme). The text has been changed to make it clear that these values are not observed values, since there are none near Darwin for this field campaign, but rather the default values used by FSBM for a maritime air mass.*

Changes to manuscript: *We have changed "in the boundary layer" to "near the surface" on P7 L8-10 to note that this initial condition of CCN does not depend on boundary layer height.*

8. Comment: P7, L9, "(not shown)" Can the authors show it, perhaps in the supplemental material. This is import if we want to use one case to represent all simulations.

Changes to manuscript: *Added Figures S2-S4 in supplemental material showing differences in TWC and MMD as a function of w, TWC, and T for the other simulated events using the Morrison and Thompson scheme. See response to Major Point #5 for more information.*

9. Comment: P7, L15: The last sentence needs to be broken into two. The sentence has two unrelated issues about inner domain and total simulation time, if I understood it correctly.

Changes to manuscript: *The referenced sentence was broken into 2 and moved to the middle of the paragraph on P7 L22-24.*

5   **10. Comment:** P10, last paragraph: The discussion of Fig. 10 needs clarification. I guess the first question I have is: why use 90%? Would 50% do? Why or why not? Also, the descriptions are hard to comprehend. Please make an effort to clarify it.

  Response: *We have attempted to clarify the meaning and utility of the minimum 90% MD in the current study.*
10 *The mass placed in the upper percentiles of the mass-size distribution (e.g. 90%) is typically from larger particles that may have been less frequently sampled by the aircraft due to the aforementioned sampling bias. However, the minimum 90% MD observed should not be impacted by this bias since it will be associated with a lack of large, dense particles and a lack of relatively high reflectivity or lightning. Therefore, if the simulation is unable to produce a single 90% MD that is as small as the observed minimum 90% MD for a given temperature, TWC, and*
15 *vertical velocity, then we feel confident in declaring a model bias. This is particularly true because the simulations produce on average ~10^2 more samples for a given TWC-T or w-T bin compared to observations. So, despite 2 orders of magnitude larger sample sizes in a simulation, the simulation is unable to produce a <u>single</u> 90% MD as small as the observed minimum 90% MD in many TWC-T and w-T bins.*

20 **Changes to manuscript**: *Clarification of the 90% MD discussion has been made on P12 L10-19.*

  11. Comment: P13, L30: "The majority of graupel at T>8C is formed by freezing raindrops". Can you give references and/or supporting evidences? My understanding is that this is only true when updraft velocity is high. Otherwise riming could be the dominant process.
25

  Response: *You are correct that the dominant process is riming. We should have stated more clearly that the primary contributor to graupel production was the heterogenous freezing of raindrops due to the collision of rain and cloud ice, which is a riming process that involves freezing of lofted raindrops. This was analyzed in the Thompson scheme and not in the FSBM scheme, however the raindrop sized MMDs and large LWCs between 0*
30 *and -4 °C in updrafts in Figure 14 indicate that the majority of the LWC is constituted by raindrops. In FSBM, by -8 °C, nearly all of the liquid is gone and most of the IWC is constituted by graupel (compare Figure 13 to Figure 12), which only decreases with decreasing temperature. This suggests that most of the graupel is being formed by raindrops that are freezing, likely heterogeneously through interactions with ice, as in the Thompson scheme. Since raindrops are smaller for T > -8 °C in FSBM, it intuitively makes sense that this contributes to smaller graupel*
35 *sizes.*

  Changes to manuscript: *Language has been changed on P15 L21-28 to clarify this and suggest this as a possible mechanism responsible for smaller graupel sizes rather than definitively concluding that it is the reason.*

40 **12. Comment:** P15, L1: "distribution tails" could mean tails at both small and large size end. May be change it to "large size tails"?

Changes to manuscript: *Added "at larger diameters" after "distribution tails" to clarify on P16 L25.*

13. Comment: Fig. 11 to 16: Can the x-axis be extended beyond 25 m/s to include higher w simulated in the model? A line can be added to indicate the observation range of 25m/s. This will give a full picture and help the readers better understand model differences and their related processes discussed in the paper.

Response: *The primary purpose of this study is to compare simulations with observations in order to establish potential model biases. Because observed vertical velocities do not exceed 25 m s^{-1}, extending the x-axis would only allow comparison between different microphysics schemes, which is not a first-order objective in this study. The evaluation of differences between schemes in Figures 11-14 is primarily to provide possible explanations for biases with respect to observations. A more in-depth analysis of differences between simulations would be interesting, but we feel that it would extend the paper beyond a reasonable length for a single study.*

Reviewer 2

Major Points

1. Comment: A table showing the density and mass-size relationships of all hydrometeor species should be provided for FSBM. I would also suggest putting terminal velocity and mass relationships to all three tables. As mentioned in the manuscript, sedimentation process also causes the different behaviors of the overestimated hydrometeor size. Some related studies (terminal velocity impact on cloud and precipitation structure) should be cited too.

Response: *Table 3 has been added to list the mass-size relationship parameters and terminal velocity-size relationship for FSBM. Terminal velocity-size relationship parameters have also been added for Thompson and Morrison in Tables 1 and 2. Note that the terminal velocity-size relationships given for FSBM in Table 3 are listed as ranges. This is because the v-D relationships used in the study come directly from ASCII files located in the /WRFV3/run/ directory of V3.6.1 of WRF. Figure R3 shows mass, density, and terminal velocity as a function of diameter in panels (a)-(c), respectively, using the ASCII file data for each of the 33 size bins of each hydrometeor species. Updated analytical formulas are not provided in the literature for the relationships shown in Figure R3, and therefore, v-D relationship as well as the ρ-D relationship for snow are listed as ranges rather than a formula.*

Changes to manuscript: *In addition to the inclusion of Table 3 and the inclusion of v-D relationship parameters for the Morrison and Thompson schemes, we have added language on P16 L4-8 to indicate that v-D relationships may impact model size biases presented in this study, but that analyzing this possible contribution is beyond the scope of the study. Relevant references have also been cited that examine impacts of sedimentation on cloud and precipitation structure.*
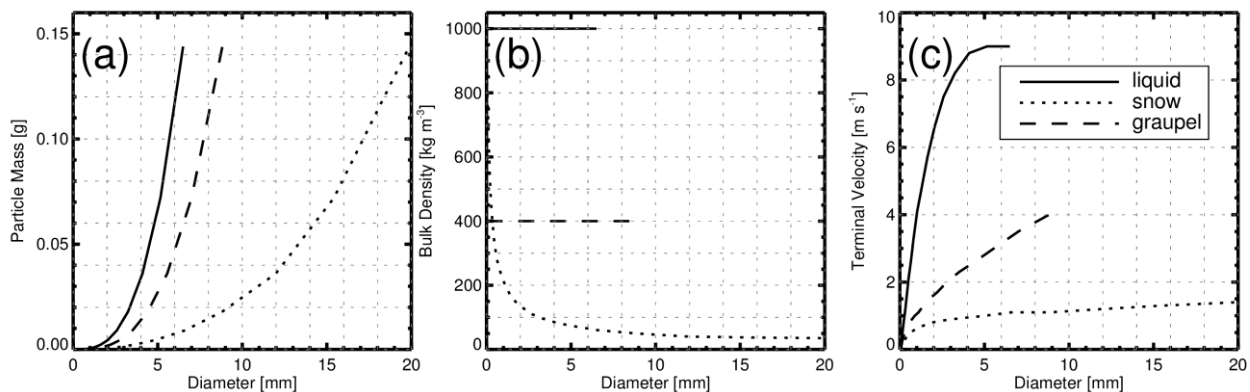

Figure R3. FSBM (a) mass , (b) density, and (c) terminal velocity as functions of diameter for liquid (solid), snow (dotted), and graupel (dashed).

2. Comment: The upper limit of the Deq of hydrometeors in FSBM is less than 10 mm due to its relatively fewer bin numbers (33 vs. 36 from many other bin schemes). How come the PSD, MSD and ZSD in Fig. 17 showing sizes reaching 10 mm for FSBM?

Response: *The upper diameter limit for snow in the FSBM scheme within WRF is actually 19.877 mm (9.9387 mm radius). This can be found by looking in the WRFV3/run/ directory for V3.6.1 in the ASCII file named "bulkradii.asc_s_0_03_0_9", which lists the corresponding bulk radius bin for each FSBM species mass bin.*
5 *Although these snow particles are extremely large, aggregates in reality achieve these sizes, and the density is assumed to be low (35 kg m⁻³).*

3. Comment: Is the C-POL reflectivity data gridded? It will be tricky to show plan view just at 2.5 km of the C-POL reflectivity in Fig. 3(a) if the data is not gridded. What is the data sample size of the observed reflectivity for
10 each level shown on Fig. 5? The unevenness of the sample size above and away from the radar can cause bias in the observed profile. How was the reflectivity calculated for each simulation? What wavelength was assumed in the model reflectivity calculation? Were mass-size relationships associated with particular schemes or the water equivalent diameter used to calculate the model reflectivity? How are the partially melted particles coated with water are treated in the calculation? Different ways of calculating the reflectivity will provide very different results.
15 Some discussions about the uncertainties of the model reflectivity should be provided.

Response: *Yes, the reflectivity data from C-POL is Cartesian gridded. Figure 5 has been revised according to Specific Comment #6 by Reviewer 1. We have provided sample sizes, normalized by domain area, of the data used in Figure 5 for each time period shown. While the C-POL domain is smaller than the WRF domains, the number
20 of grid points ≥ 5 dBZ per km² is comparable between C-POL and WRF for the 12Z-18Z time period, while the C-POL sample size per km² is larger than WRF for the 18Z-00Z time period. Additionally, the larger sample sizes used by analyzing 6-hr time periods (rather than a single time step that was previously shown) suggests that sample size limitations would not erase the extremely large difference in reflectivity aloft shown in the 99ᵗʰ percentile profiles of Figure 5.*
25

*Reflectivity calculations assume Rayleigh scattering, which is valid for comparison with the (5.5-cm) C-POL wavelength in this environment. This calculation follows Smith (1984), is commonly used in studies, and is consistent with PSDs and single particle properties assumed in each of the schemes. Rayleigh reflectivity is now output by default for several microphysics schemes in WRF including the Thompson and Morrison schemes used
30 here. As discussed in Smith (1984), the estimated Rayleigh reflectivity for ice could be slightly off, but any potential error is much smaller than the very large differences between observations and simulations shown in Figure 5. The equation does indeed account for the m-D relationship via the melted equivalent diameter calculation (Equation 5), and thus the reflectivity is diagnosed to change with changes in the m-D relationship.*

35 *Partially melted particles coated with water are not explicitly represented in the microphysics schemes. The peak in reflectivity near the melting level in the bulk schemes shown in Figure 5 is the result of a very simple parameterization in the WRF computed reflectivity, which is implemented to represent the "bright band" caused by water coated melting ice particles. This peak does not occur for FSBM since FSBM reflectivity is computed offline by summing reflectivity contributions from each hydrometeor size bin without implementation of the melting
40 layer algorithm. Although the reflectivity peak caused by water coated melting ice could have significant errors, this does not impact our conclusions drawn for regions above this melting layer, which is the region we focus on.*

Changes to manuscript: *Importantly, we note that we have used an updated C-POL dataset provided by Alain Protat who has been added as a co-author. We have also gridded these C-POL radial sweep files ourselves to match the simulation horizontal grid spacing (1 km) and found a ~2.5 dB increase, which is why the cross-section figures (Figures 3 and 4) change from the original version, however this increase is not central to any of our conclusions.*

We have included statements that C-POL data is Cartesian gridded and uses 1-km horizontal and 500-m vertical grid spacing in the manuscript on P9 L16-18. Figure 5 has been revised, as discussed in the response above and in Specific Comment #6 by Reviewer 1, and relevant discussion of Figure 5 has been revised on P10 L4-15

We have moved Equations 5-7 from Section 5.5 to the beginning of Section 5.1 (P9) to introduce the calculation of reflectivity when the subject is initially discussed.

We state that simulated reflectivity is calculated assuming Rayleigh scattering on P9 L14-15 and provide a brief explanation of why this is a good assumption.

Language has been added on P9 L9-13 to note the differences in reflectivity calculations between FSBM and bulk schemes and to note that the bulk schemes use a parameterization to treat reflectivity calculations of partially melted ice coated with water but FSBM does not.

4. Comment: Figure 4 is the same as Fig. 3 and Fig. 9 is the same as Fig. 8. It is hard to follow the discussions about these figures.

Changes to manuscript: *The correct version of Figure 4 is now in place. The correct version of former Figure 9 is now represented in the bottom panel of Figure 7 (g-i), by which the old Figures 7-9 are now combined into a 3x3 panel that is the current Figure 7, as recommended in Specific Comment #2 by Reviewer 1.*

5. Comment: It will be good to plot PSD of liquid and ice particles separately in Figs. 11 and 12.

Response: *While separating these figures into liquid and ice is insightful from a model only perspective, we use the combined hydrometeor distributions because observations are not separated into liquid and ice species. Simulated liquid and ice MDs as a function of vertical velocity are shown in Figure R4 for T between -32 °C and -40 °C and Figure R5 for T between -8 °C and -16 °C, however they are difficult to interpret without the contribution of each species to the total mass for a given vertical velocity (e.g., liquid MDs can be very small, but not contribute much to the overall MDs if liquid doesn't contribute much to the overall mass). This is the reason that we feel that it is better to draw inferences from Figures 11-14 in explaining some differences in (current) Figures 7-10. For example, Figure 14 shows that for temperatures between -8 °C and -16 °C, LWC can be significant in the bulk schemes for $w > 5$ m s^{-1} with corresponding very small MMDs, whereas LWC is insignificant in FSBM and MMDs are significantly larger. Snow and graupel MMDs in Figures 12-13 are large in all schemes for these temperatures, and therefore, it is known that cloud droplets must be causing the sharp decrease in 10% MD with increasing vertical velocity in bulk schemes in Figure 10.*
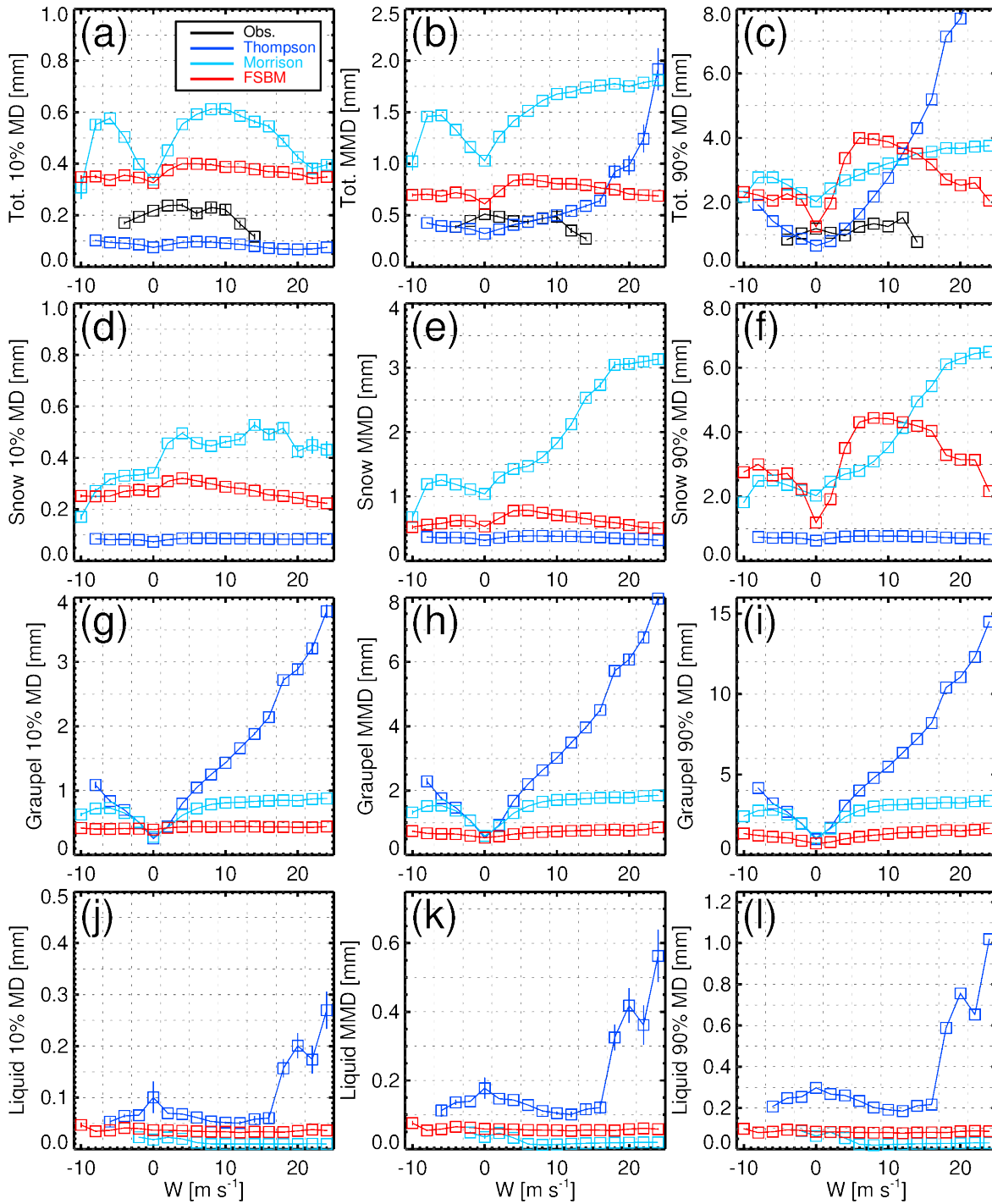
Figure R4. (a)-(c) Combined hydrometeor 10% MD, MMD, and 90% MD, respectively, as a function of *w* for T between -32 °C and -40 °C. (d)-(f) As in (a)-(c), but for snow MDs. (g)-(i) As in (a)-(c), but for graupel MDs. (j)-(l) As in (a)-(c), but for liquid MDs.
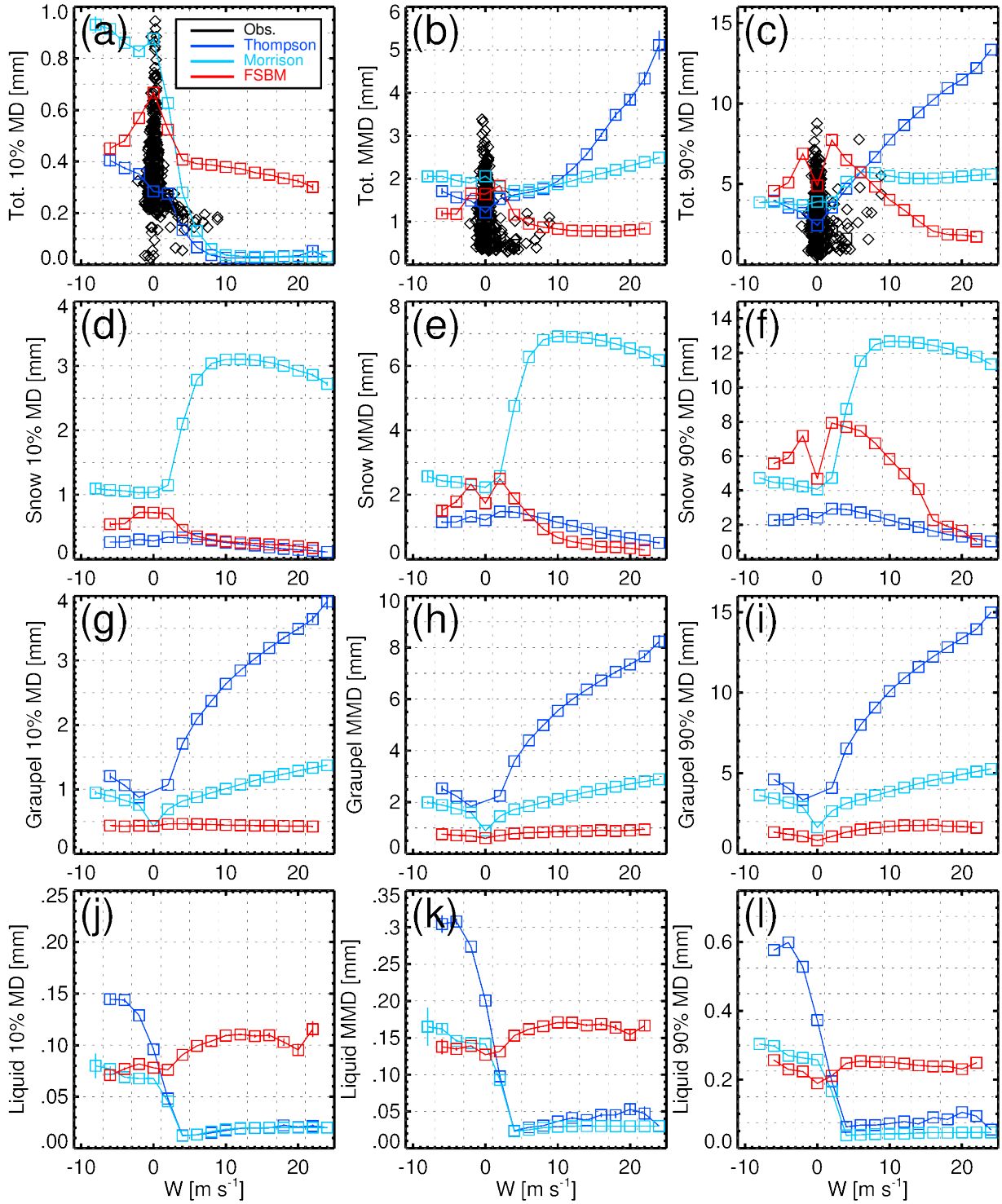
Figure R5. As in Figure R4, but for a temperature range between -8 °C and -16 °C.

6. Comment: The last sentence in the abstract is too strong. It is hard to imagine current microphysics schemes will uniformly produce a high bias in reflectivity.

Changes to Manuscript: *Changed language on P1 L26-28 to indicate that this statement refers only to the microphysics schemes evaluated in the current study and not all microphysics schemes in general.*

Minor Points

1. Comment: It will be good to also mention the lower size limit of the OAPs in page 4 line 15.

Changes to manuscript: *Added lower limits of OAPs to P4 L16-17.*

2. Comment: The threshold for simulated condensation mass mixing ratio is too small at 10-12 kg kg-1. Using this threshold may introduce grid points with unrealistic results. 10-6 kg kg-1 should be a good threshold for the analysis.

Changes to manuscript: *All simulated and observed data now use a constraint of > 0.1 g m^{-3}, which is the approximate TWC observational uncertainty at -40°C (see discussion in manuscript on P4 L29-33.*

3. Comment: You probably referred to Fig. 6b in line 14 on page 12.

Changes to manuscript: *Changed reference from Fig. 6c to Fig. 6b on P14 L6.*

4. Comment: Page 1, line 16: using "different" microphysics. . .

Changes to manuscript: *Changed "differing" to "different" on P1 L16.*

A ubiquitous ice size bias in simulations of tropical deep convection

McKenna W. Stanford[1], Adam Varble[1], Ed Zipser[1], J. Walter Strapp[2], Delphine Leroy[3], Alfons Schwarzenboeck[3], ~~and~~ Rodney Potts[4], and Alain Protat[4]

[1] Department of Atmospheric Sciences, University of Utah, Salt Lake City, Utah, 84112, USA
[2] Met Analytics, Inc., Aurora, Ontario, Canada
[3] Université Clermont Auvergne/CNRS, Laboratoire de Météorologie Physique, Clermont-Ferrand, France
[4] Australian Bureau of Meteorology, Research and Development Branch~~Center for Australian Weather and Climate Research~~, Melbourne, Victoria, Australia

Correspondence to: McKenna W. Stanford (mckenna.stanford@utah.edu)

Abstract. The High Altitude Ice Crystals – High Ice Water Content (HAIC-HIWC) joint field campaign produced aircraft retrievals of total condensed water content (TWC), hydrometeor particle size distributions (PSDs), and vertical velocity (*w*) in high ice water content regions of mature and decaying tropical mesoscale convective systems (MCSs). The resulting dataset is used here to explore causes of the commonly documented high bias in radar reflectivity within cloud-resolving simulations of deep convection. This bias has been linked to overly strong simulated convective updrafts lofting excessive condensate mass but is also modulated by parameterizations of hydrometeor size distributions, single particle properties, species separation, and microphysical processes. Observations are compared with three Weather Research and Forecasting model simulations of an observed MCS using differ~~ing~~ent microphysics parameterizations while controlling for *w*, TWC, and temperature. Two popular bulk microphysics schemes (Thompson and Morrison) and one bin microphysics scheme (Fast Spectral Bin Microphysics) are compared. For temperatures between -10 °C and -40 °C and TWC > 1 g m^{-3} ~~inside updrafts~~, all microphysics schemes produce median mass diameters (MMDs) that are generally larger than observed, and the precipitating ice species that controls this size bias varies by scheme, temperature, and *w*. Despite a much greater number of samples, all simulations fail to reproduce observed high TWC conditions (> 2 g m^{-3}) between -20 °C and -40 °C in which only a small fraction of condensate mass is found in relatively large particle sizes greater than 1 mm in diameter. Although more mass is distributed to relatively large particle sizes relative to observed across all schemes when controlling for temperature, *w,* and TWC, differences with observations are significantly variable between the schemes tested. ~~with observations for a given particle size vary greatly between schemes.~~ As a result, this bias is hypothesized to partly result from errors in parameterized hydrometeor PSD and single particle properties, but because it is present in all schemes, it may also partly result from errors in parameterized microphysical processes present in all schemes. Because of these ubiquitous ice size biases, the frequently used microphysical parameterizations evaluated in this study inherently produce a high bias in convective reflectivity for a wide range of temperatures, vertical velocities, and TWCs.

1 Introduction

Improving parameterizations is inherently challenging because they are often guided by quite crude observational constraints that fail to cover the large range of atmospheric conditions and cloud responses possible (Tao and Moncrieff, 2009; Khain et al., 2015). A common method used to evaluate the wide range of model physics schemes available is model intercomparison,
5 which involves rigorous comparison of cloud resolving model (CRM), limited area model (LAM), single column model (SCM), or general circulation model (GCM) output with high-quality in situ and remote-sensing observations, often from field experiments (e.g. Bechtold et al., 2000; Matsui et al., 2009; Wang et al., 2009a,b; Fridlind et al., 2010; Varble et al., 2011; Zhu et al., 2012; Varble et al., 2014a,b; and many others). Many resulting studies have focused on the characterization of cloud microphysical and dynamical properties in tropical deep convective systems (e.g. Stith et al., 2002, 2004; Heymsfield, 2003a,b;
10 Heymsfield et al., 2005, 2009, 2010; Lawson et al., 2015), which are climatically important because of their large contribution to global precipitation (Nesbitt et al. 2006). Ice microphysical properties are of particular interest because of anvil cirrus radiative effects (Hartmann et al., 1992; , Heymsfield and McFarquhar, 1996; McFarquhar and Heymsfield, 1996, 1997; Garrett et al., 2005) and the strong sensitivity of simulated deep convective systems to the parameterizations of species, habit, size distributions, and process rates (Chen and Cotton, 1988; McCumber et al., 1991; Gilmore et al., 2004; Milbrandt and Yau,
15 2005a; McFarquhar et al., 2006; Morrison and Grabowski, 2008).

This study utilizes data collected from the High Altitude Ice Crystals – High Ice Water Content (HAIC-HIWC) joint field campaign (Dezitter et al., 2013; Strapp et al., 2016). The first HAIC-HIWC phase was conducted in and around Darwin, Australia, from January to March in 2014 during the wet monsoon season. HAIC-HIWC was primarily designed to investigate meteorological processes responsible for commercial aircraft engine malfunction hypothesized to result from ingestion of high
20 ice water content (> 2 g m^{-3}) in regions of low radar reflectivity that did not raise alarm to pilots (Lawson et al., 1998, Mason et al. 2006). The campaign aircraft was thus equipped with instruments that measured particle cross-sectional areas from which particle size distributions (PSDs) were retrieved and total water content (TWC) from which condensate mass was retrieved in and around convective regions of tropical mesoscale convective systems (MCSs) where high ice water content (IWC) was expected to be encountered (Leroy et al., 2016a, hereafter referenced as L16a).

25 The HAIC-HIWC datasets provide a wealth of in-cloud microphysical measurements that are ideal for evaluating simulations of tropical deep convective systems. One commonly documented model bias is the overestimation of radar reflectivity aloft in tropical deep convection, which has been shown to result from excessive graupel production in simulations (Blossey et al., 2007; Lang et al., 2007; Li et al., 2008, Matsui et al., 2009; Varble et al., 2011; Caine et al., 2013), but also from overly large snow sizes in some two-moment bulk microphysics schemes (Varble et al., 2011). Varble et al. (2014a)
30 explored the possible contribution of vertical velocity to this bias, and concluded that overly strong convective updrafts lofting large amounts of condensate mass in simulations were partially responsible for the reflectivity bias. However, they also found that the magnitude of this bias depends on the parameterization of microphysics. Franklin et al. (2016) compared a single-moment bulk microphysics scheme with HAIC-HIWC data and found that ice particle size aloft strongly impacts updraft

buoyancy and radar reflectivities in simulations of tropical deep convection, signaling the importance of interacting microphysics and dynamics. However, microphysical parameterization and convective dynamics contributions to the reflectivity bias have yet to be untangled, and individual contributions of hydrometeor type, size, and bulk mass to the bias have yet to be separated.

5   Lang et al. (2011, 2014) improved upon a single-moment bulk microphysics scheme by adjusting model physics for the purpose of reproducing observed reflectivity statistical distributions. However, their study was limited to using remote-sensing data, which provides limited information about vertical velocity and hydrometeor size, species, and bulk mass because of significant retrieval assumptions and uncertainties. In situ retrievals of vertical velocity, ice crystal size, and bulk mass during HAIC-HIWC provide a dataset needed to explore the various interacting contributors to model reflectivity biases. In

10 particular, mass-size distributions (MSDs) constructed from these measurements may be compared with simulated MSDs in microphysics schemes. L16a and Leroy et al. (20176b) (hereafter referenced as L176b) perform extensive analyses on median mass diameters (MMDs) retrieved from HAIC-HIWC data, a characteristic size metric that provides the diameter at which half of the bulk mass resides in smaller diameters and half resides in larger diameters. By comparing hydrometeor sizes and investigating the model parameters and processes that control size distributions, possible sources for biases may be identified

15 and further observational constraints on these parameters may be implemented.

        This study focuses on comparing simulated and observed hydrometeor sizes (e.g., MMDs) as functions of bulk mass and vertical velocity so that the role of microphysical processes and assumed particle properties in producing model convective precipitation biases can be isolated from the roles of total condensate and vertical velocity. Two bulk microphysics schemes and one explicit bin microphysics scheme are evaluated, providing insight into how the bias and its causes differ between

20 fundamentally different approaches to microphysics parameterization. The extent to which a reflectivity bias exists in bin microphysics schemes has not been extensively explored, although Ackerman et al. (2015) show that a CRM simulation using bin microphysics failed to reproduce observed low reflectivity values in high IWC regions, suggesting that this bias may exist across both bulk and bin schemes. Observations are described in Sect. 2, model setup and microphysics schemes in Sect. 3, intercomparison methodology and limitations in Sect. 4, results in Sect. 5, and conclusions in Sect. 6.

25 **2 Observations**

The majority of data collected during the first HAIC-HIWC field campaign was from instrumentation aboard the SAFIRE[1] Falcon 20 research aircraft. Strapp et al. (2016) describe the sampling strategy of the Falcon 20 during the HAIC-HIWC Darwin campaign, in which 17 out of 23 total flights targeted regions of large mature and decaying tropical MCSs with cold infrared (IR) brightness temperatures observed by satellite. Many flight legs penetrated convective updraft cores or regions

30 downstream of updraft cores at temperatures near -30 °C and -40 °C, with fewer flight legs performed around -50 °C and -10

---

[1] Service des Avions Francais Instrumentes pour la Recherche en Environnement

°C levels. Example flight tracks are shown in Fig. 1, overlaid on 10.8-µm IR brightness temperatures observed by the Japanese Multifunctional Transport Satellite 1R (MTSAT-1R) satellite, where the satellite image is a single time from when the system was sampled. MCSs sampled during Flight 6 on 23 January 2014 (Fig. 1a), Flights 12-13 from 2-3 February 2014 (Fig. 1b), Flight 16 on 7 February 2014 (Fig. 1c), and Flight 23 on 18 February 2014 (Fig. 1d) are shown because these are the events that were successfully simulated using the two bulk microphysics schemes. These events were chosen for simulation because of many observations at different temperature levels penetrating many convective updrafts with common occurrences of high IWC exceeding 2 g m^{-3}. Figure S1 shows the distribution of HAIC-HIWC samples by flight that are used in this study, with simulated event dates labeled. The mean and standard deviation of temperature for each flight is also shown in Figure S1.

Vertical velocities (w) were calculated by SAFIRE using a method similar to that of Jorgensen and LeMone (1989), in which w is defined as the difference between the vertical motions of the aircraft relative to the ground and relative to the air. The vertical motion with respect to air is calculated using the aircraft's true air speed along with attack, side-slip, pitch, and roll angles, the former two of which are measured using differential pressure measurements and the latter two of which use inertial navigation measurements. Uncertainty in w calculations is estimated at ~ 1 m s^{-1} (Jorgensen and LeMone, 1989).

Particle cross-section images used for derivations of PSDs were obtained by two optical array probes (OAPs): the 2D-Stereo probe (2D-S, Lawson et al., 2006) from SPEC Inc. and the Precipitation Imaging Probe (PIP, Baumgardner et al., 2011) from Droplet Measurement Technologies. The 2D-S measures particles with diameters ranging fromless than 10-1280 µm at 10 µm resolution, and the PIP measures particles ranging fromup to 100-6400 µm at 100 µm resolution. Linearly weighted composite PSDs for 5-second sampling windows were constructed using both OAPs, as described in L16a. For these PSDs, truncated particle images were reconstructed using the method of Korolev and Sussman (2000). Based on spherical assumptions, truncated PIP images included in the PSDs derived by L16a reach diameters exceeding 12 mm. The OAPs were equipped with anti-shattering tips to avoid ice fragmentation (Korolev and Isaac, 2005; Heymsfield, 2007; Lawson, 2011) and an inter-arrival time algorithm was used to remove potentially shattered particles (Field et al., 2003, 2006; Korolev and Field, 2015).

Bulk TWC measurements were made with an isokinetic evaporator probe (IKP) first designed and developed by the National Research Council of Canada (NRC) and then re-engineered as the IKP2 by Science Engineering Associates (SEA) Inc. and NRC to operate within the constraints of the Falcon 20 aircraft (Davison et al., 2008; Strapp et al., 2016; Davison et al., 2016). The probe was designed specifically for the measurement of high altitude, high IWC conditions, with a target accuracy of 20%. The IKP2 uses a differential hygrometry method in calculating condensed water content (hereafter, TWC) that accounts for the subtraction of background water vapor. However, due to contamination of the background water vapor by ice crystals during the Darwin 2014 flight campaign, it was necessary to assume ice saturation inside cloud. There is therefore a fundamental TWC uncertainty due to the difference between the actual vapor concentrations in cloud versus the ice saturation estimate that may reach several tenths of a g m^{-3} at -10 °C, dropping to about 0.1 g m^{-3} at -40 °C. Work is continuing to better quantify this uncertainty, which is of course proportionately larger for low-IWC sections of cloud. Because of this uncertainty, observed samples with TWC ≤ 0.1 g m^{-3} are excluded from this study. Further details about the

19

microphysical instrumentation used on the Falcon 20 and data processing from the OAPs and IKP2 may be found in L16a, L17~~6b~~, and Strapp et al. (2016).

L16~~a~~ use retrieved TWC with retrieved PSDs to constrain mass-size relationships ($m = \alpha D^\beta$) used to calculate MSDs over 5-second sampling intervals. The exponent β is constrained through relation to the exponents of area-size and perimeter-size power law relationships derived from OAP images, which allows it to vary as a function of time as crystal habits change during flight. The parameter α is constrained by matching the integrated MSD to TWC measurements from the IKP2. The particle diameter size definition used for most results in this study is the 2D area equivalent diameter (D_{eq}), defined as the diameter of a circle with the same area as particle images from the OAPs. However, some statistics are shown using the D_{max} definition, defined here as the maximum dimension through the center of the 2D crystal image.

Mass-size distributions, defined as the product of the PSD, given by $N(D)$, and the mass-size power law relationship, are given by Eq. (1):

$$M(D) = \alpha D^\beta N(D) \tag{1}$$

where $M(D)$ has units of kg m⁻⁴. Percentiles of the MSD are used to describe hydrometeor size, and are calculated by numerically integrating the MSD from 0 to the mass diameter (MD) where the integrated mass equals the desired percentage of total mass. In particular, the MMD is used for comparison of retrieved and simulated hydrometeor sizes, where the MMD is calculated using Eq. (2):

$$\int_0^{MMD} \alpha D^\beta N(D) dD = \frac{1}{2}\int_0^\infty \alpha D^\beta N(D) dD = \frac{1}{2}TWC \tag{2}$$

Importantly, we note that large uncertainties exist in retrievals of total number concentration for diameters smaller than 150 μm based on the derivation technique employed (not shown). However, TWC and MSD measurements that are based on higher moments of the PSD are more certain since the IKP2 measures TWC plus water vapor, where the subtraction of water vapor is the primary source of uncertainty, and MSDs are not strongly impacted by particles smaller than 150 μm (L17). Moreover, t~~T~~he ~~PSD and~~ MSD dataset~~s~~ derived by L16~~a~~ permit~~s~~ comparison~~s~~ of simulated and observed hydrometeor properties in the context of TWC and w that are more accurate and detailed than remote-sensing retrievals. However, because an objective of this study is to investigate well-known reflectivity biases, data from a C-band scanning dual-polarimetric radar (C-POL) (Keenan et al., 1998) located near Darwin are utilized for Flight 23 on 18 February 2014, one of the only events that occurred within range of the radar. The C-POL dataset for 18 February permits establishing reflectivity biases in the current study and provides motivation for focusing on this particular simulated event for comparison with observed microphysical quantities.

3 Model

3.1 Model Setup

The Advanced Research Weather Research and Forecasting (WRF-ARW) V3.6.1 model (Skamarock et al., 2008) is used to simulate the MCS events shown in Fig. 1 using two different bulk microphysics schemes, while the 18 February event is also simulated using a bin microphysics scheme. The Bureau of Meteorology's (BoOM) Australian Community Climate and Earth-System Simulator Regional model (ACCESS-R) (Puri et al., 2013) analyses with 12-km horizontal grid spacing are used for initial and boundary conditions~~large scale forcing~~. Physics parameterizations common to all simulations performed include the Mellor-Yamada-Janjic (MYJ) planetary boundary layer (PBL) scheme (Janjic, 1994), the Rapid Radiation Transfer Model (RRTM) longwave radiation scheme (Mlawer et al,. 1997), the Dudhia (1989) shortwave radiation scheme, the Kain-Fritsch cumulus scheme (Kain, 2004), and the Noah Land Surface Model (Chen and Dudhia, 2001). All of the simulated MCS events use 9:3:1-km two-way nesting with 92 vertical levels and the~~an~~ innermost 1000-m grid spacing domain is used for analysis. The nested domains and C-POL coverage for the 18 February simulations are shown in Fig. 2.

3.2 Microphysics

The microphysics schemes employed in this study include the Thompson (Thompson et al., 2008) and Morrison (Morrison et al., 2009) bulk microphysics schemes and the Hebrew University Fast Spectral Bin Microphysics (FSBM) scheme (Lynn et al., 2005). The Thompson and Morrison bulk schemes predict integral moments of the PSD for five hydrometeor species, including cloud ice, cloud water, rain, snow, and graupel. PSDs in both bulk schemes are generally represented by a gamma function of the following form:

$$N(D) = N_0 D^\mu e^{-\lambda D} \tag{3}$$

where N_0 is the intercept parameter, D is the particle diameter, μ is the shape parameter, and λ is the slope parameter. N_0 essentially controls the number of small particles, μ controls the dispersion of the PSD, and λ controls the slope of the PSD. Snow, graupel, rain, and cloud ice are double-moment species in the Morrison scheme with both mass mixing ratio (q) and number concentration (N) predicted, whereas cloud water is a single-moment species (prognostic q only and constant N). The Thompson scheme uses double-moment rain and cloud ice with single-moment snow, graupel, and cloud water. An exception to the generalized gamma distribution given by Eq. (3) is the Thompson snow PSD parameterization, which uses a bimodal gamma distribution given by Field et al. (2005) that varies with temperature. The Morrison scheme assumes a constant bulk density for all ice species given by Reisner et al. (1998), while Thompson only assumes constant graupel and cloud ice density. For snow, the mass-size power law relationship presented in Cox (1988) is used that allows the bulk density of snow to vary with particle size. Varble et al. (2014a) showed that this relationship (where $m \propto D^2$) reproduces observed reflectivity better than schemes assuming $m \propto D^3$ for snow, while others have shown that it is supported by surface disdrometer (Mitchell et al., 1990) and aircraft (Westbrook et al., 2004) observations of snow particles. While the Thompson scheme uses a constant bulk density typical of medium density graupel, it varies N_0 inversely as a function of the predicted q and shifts the fall-speed

21

relationship from graupel toward hail as particle size increases. For the versions of the Morrison and Thompson schemes used in this study, $\mu = 0$ for rain, graupel, and cloud ice, making the PSD exponential. For snow, $\mu = 0$ for Morrison, but is nonzero in the Thompson scheme. Cloud ice μ is also variable in both schemes. Lastly, this study uses a constant cloud droplet number concentration of 100 cm^{-3} in both bulk schemes that is typical of the clean tropical maritime air masses commonly observed in

5   the vicinity of Darwin during the active monsoon. A summary of the PSD, and mass-size relationship, and terminal velocity-size relationship parameters used in the bulk schemes may be found in Tables 1 and 2.

The FSBM scheme uses 33 mass-doubling bins to represent PSDs, and process rates are computed separately for each bin. FSBM species include cloud condensation nuclei (CCN), liquid water (both raindrops and cloud droplets), graupel, and ice crystals/aggregates (hereafter referred to as snow). Graupel has a bulk density of 400 kg m^{-3}, and liquid has a bulk density

10   of 1000 kg m^{-3}. The density of vapor-grown ice in FSBM varies from 900 kg m^{-3} at small sizes to 35 kg m^{-3} at large sizes. It also explicitly represents cloud droplet nucleation whereas the bulk schemes do not, and it is able to maintain supersaturations over liquid that the bulk schemes cannot. Initial CCN concentrations in the FSBM scheme are also set to resemble the maritime environment in Darwin using FSBM's default maritime airmass aerosol concentration, in which, where concentrations near the surface concentrations in the boundary layer are ~ 100 cm^{-3} and decrease exponentially with height to ~ 50 cm^{-3} at 4-km

15   altitude and < 10 cm^{-3} at 9-km altitude. A summary of FSBM mass-size and terminal velocity-size relationships are given in Table 3.

4 Intercomparison Methodology

Simulating every HAIC-HIWC event with multiple model setups is not computationally feasible, and therefore, four MCS events shown in Fig. 1 are simulated using the Morrison and Thompson microphysics schemes. The primary differences

20   between simulated events are the mesoscale precipitation structure and peak convective intensities, while simulated hydrometeor properties vary little between events when controlling for w and TWC (see Section 5.2.2 and respective discussion of Figures S2-S4 not shown). Much larger differences exist between simulations with different microphysics schemes when simulating the same case, suggesting that a single simulated event is adequate for robustly examining differences between various microphysics schemes and observations. Moreover, Figure S1 shows that using observations from the entire Darwin

25   field campaign is necessary in order to stratify MMD and TWC by temperature with sufficient sample sizes. The 18 February event (Flight 23) contains the most observations in high TWC conditions near -10 to -15 °C and is the only flight within observing range of C-POL, two primary reasons that simulations of it are used for comparison with observations. The 450-km by 540-km inner domain (1000-m grid spacing) for this case is shown in Fig. 2. The simulation was run for 30 hours from 00Z on the 18th to 06Z on the 19th. Observations from all HAIC-HIWC flights are compared with results from the innermost domain

30   the simulated 18 February 2014 MCS (Flight 23) for a 6-hour period between 18Z on the 18th and to 00Z on the 19th during the mature and decaying stages of the MCS. . This event contains the most flight observations near -10/-15 °C and flight legs within range of C-POL. The 450-km by 540-km inner domain (1000-m grid spacing) for this case is shown in Fig. 2 with model output from 00Z on the 18th to 06Z on the 19th.

Four primary variables are analyzed: temperature (T), TWC, w, and percentiles of the MSD (i.e., 10% MD, MMD, and 90% MD). Observed MDs and bulk mass are not separated by species. However, L16 state that only trace amounts of liquid water content (LWC) were detected for a few flights at T > -20 °C, and thus, TWC is a reasonable proxy for IWC in the vast majority of observed situations. Relatively small amounts of LWC measured in the mature and decaying MCSs sampled during HAIC-HIWC differ from measurements in isolated, growing cumulus cells such as those sampled during the Ice in Clouds Experiment – Tropical field campaign, which show a considerable amount of supercooled liquid at temperatures down to -15 °C (Heymsfield and Willis, 2014; Lawson et al., 2015; Yang et al., 2016). Simulated bulk mass and MDs are calculated for both individual and combined species, where the combined-hydrometeor MSD, given by $M(D)_{tot}$, is the composite MSD consisting of all hydrometeor species in the scheme, as given by Eq. (4):

$$M(D)_{tot} = \sum_{i=1}^{n} \alpha_i D^{\beta_i} N_i(D) \tag{4}$$

where n is the number of species in the microphysics scheme. For the evaluation of mass partitioning between species in bulk schemes, liquid MSDs combine cloud water and rain while snow MSDs consist of all vapor grown ice: cloud ice and snow.

Comparison of simulation output and measurements are ~~primarily confined to "updraft"~~confined to flight segments and grid points containing TWC > 0.1 g m^{-3} in order to avoid large observational uncertainty associated with smaller TWC values~~with (1) $w \geq 1$ m s^{-1} and (2) condensate mass mixing ratio > 10^{-12} kg kg^{-1}~~. Comparisons are also limited to -60 °C ≤ T ≤ 0 °C. With these constraints, grid point sample sizes for individual simulated events are greater than 10^6, approximately 2~~3~~ orders of magnitude greater than the observational sample size. While TWC and MDs are analyzed for both positive and negative vertical velocities, emphasis is placed on updrafts since the production of high IWC requires condensation of a substantial amount of water vapor, whereas adiabatic warming and drying in downdrafts typically counteracts this process and leads to evaporation. However, since regions of high IWC certainly exist outside of updrafts after they are detrained, downdrafts are included to determine possible ice size biases in both types of convective motions. Along with the bulk mass constraint, "updrafts" are defined as points with $w > 1$ m s^{-1}, "downdrafts" as points with $w < 1$ m s^{-1}, and points with w between -1 and 1 m s^{-1} are considered relatively quiescent regions.

A possible comparison bias may result from different grid spacing in simulations and observations. Simulated events are analyzed within the 1000-m horizontal grid spacing domain, whereas observed PSDs are retrieved using 5-second sampling windows, which corresponds to a grid spacing of ~ 750 m assuming a typical aircraft speed of 150 m s^{-1}. However, similar results between a 333-m grid spacing domain simulation and the 1000-m domain for the 18 February event (not shown) suggest that this grid spacing difference does not significantly contribute to large differences between simulations and observations.

The most significant source of comparison bias is the subjective observational sampling. Regions with lightning or "red" on the pilot's X-band radar display (reflectivity exceeding 40 dBZ) were avoided during flights, and these regions likely contain the most intense convective cells with the most graupel and liquid water (e.g., Zipser and Lutz, 1994). This sampling cannot be replicated in simulations because of the previously mentioned biases in simulated reflectivity and the lack of simulated lightning. Reflectivities exceeding 40 dBZ were infrequent at flight level for most flights, with the exception of legs

at temperatures near -10 °C, however lightning was common in the most intense convective cells a~~a~~ssociated with many of the large MCSs. Possible effects of this bias on interpretation of results are discussed further in subsequent sections.

5 Results

5.1 Radar reflectivity

5   Simulated Rayleigh radar reflectivity (Z_e) for bulk schemes is calculated by integrating over the sixth moment of the melted equivalent diameter ($D_{eq,melt}$) size distribution for each individual hydrometeor species and summing Z_e for all species. For ice particles, $D_{eq,melt}$ is given by Eq. (5):

$$D_{eq,melt} = \left[\frac{6\alpha}{\pi\rho_w}\right]^{\frac{1}{3}} D^{\frac{\beta}{3}} \qquad (5)$$

where ρ_w is the bulk density of water. Reflectivity for bulk scheme ice hydrometeors is then given by Eq. (6):

$$Z_e = 0.224\times10^{18} \int_0^\infty D_{eq,melt}^6 \, N\big(D\{D_{eq,melt}\}\big) \left|\frac{\partial D}{\partial D_{eq,melt}}\right| dD \qquad (6)$$

where 10^{18} is a conversion factor from m^6 to mm^6 and 0.224 is a factor accounting for the different dielectric constants of ice
10   and liquid, following Smith (1984). For liquid water species in bulk schemes, $D_{eq,melt} = D$ and Eq. (6) reduces to Eq. (7):

$$Z_e = 10^{18} \int_0^\infty D^6 N(D) dD \qquad (7)$$

Recall that the FSBM scheme does not assume a continuous PSD but rather computes number concentrations for discrete particle size (and mass) bins. Therefore, FSBM reflectivity is calculated for each discrete bin using the integrands of Eq. (6) and Eq. (7) and then summed to give a total reflectivity. Another important difference between reflectivity calculations in
15   FSBM and bulk schemes is the use of a parameterization in bulk schemes to account for partially melted ice coated with water (Blahak, 2007) which is not used in FSBM, however this does not impact our analyses since they focus on sub-freezing temperatures. The presented (Rayleigh) approximation for simulated reflectivity works well for comparison with C-POL's wavelength (5.5-cm) in convective systems without a significant amount of large hail as is expected during Darwin's active monsoon period. C-POL reflectivity is interpolated to a Cartesian grid with a 1-km horizontal grid spacing and 500-m vertical
20   grid spacing. Simulated reflectivity used for comparison maintains the native 1-km horizontal grid spacing but is interpolated vertically to C-POL's constant altitude levels for comparisons.

Observed and simulated Rayleigh radar reflectivity horizontal cross-sections are shown for the 18 February 2014 MCS at 18Z in Fig. 3-4 for 2.5-km (~ 13 °C) and 7-km (~ -10 °C) altitudes, respectively. Observed 2.5-km altitude reflectivities are typically 40-45 ~~- 40~~ dBZ in convective cores with an expansive stratiform region containing reflectivities between 25-35 ~~35~~
25   dBZ. Both bulk schemes produce much more widespread regions of reflectivity exceeding 40 dBZ with convective core values reaching 55 dBZ and more convective organization than observed. The FSBM scheme produces reflectivity values more

closely to those observed, although the system is somewhat less organized than observed and produces lesser maximum reflectivities than C-POL at this altitude. At 7-km (Fig. 4), observed reflectivities mostly remain below 25~~0~~ dBZ, with values reaching 30~~25~~-35 dBZ for only the most intense reflectivity cores. Conversely, all simulated reflectivities at 7-km exceed 25 dBZ in much more expansive regions compared to C-POL. Thompson produces a cellular structure with cores exceeding 55

5   dBZ, and Morrison and FSBM produce widespread regions with reflectivities > 40 dBZ. Fig. 3-4 clearly show differences in both the horizontal and vertical precipitation structures of the simulated and observed MCSs, and changing the microphysics parameterization also produces substantially different results.

More is revealed through examination of vertical reflectivity profiles. Figure 5 shows ~~90th and~~ 99th percentile profiles of reflectivity ≥ 5 dBZ for the 18 February MCS during two time periods: 12Z-18Z on the 18th (Fig. 5a) and 18Z on the 18th

10   to 00Z on the 19th (Fig. 5c). Sample sizes normalized by domain area are shown in Fig. 5b and Fig. 5d for the 12Z-18Z and 18Z-00Z time periods, respectively. Two time periods are shown because the most intense observed convection occurred before 18Z, but the flight took place after this time when the most intense convection had moved out of the C-POL domain. During the 12Z-18Z time period, all simulated 99th percentile reflectivities exceed C-POL 99th percentile reflectivities by 10-15 dBZ above the melting level throughout the majority of the free troposphere. ~~at 16Z, a time representative of the MCS~~

15   ~~upscale growth stage when observed convection was most intense, as judged by upper tropospheric reflectivity and lightning flash rate.~~ Differences between C-POL and simulations are larger in the 18Z-00Z time period where the simulated 99th percentile reflectivities exceed observed 99th percentile reflectivities by up to 20 dBZ above the melting level. T~~t~~hese large discrepancies in both time periods suggest a significant reflectivity bias at sub-freezing temperatures.~~All simulations produce~~ ~~90th percentile reflectivities that exceed C-POL reflectivities by 10-20 dBZ in the mid-upper troposphere, with less significant~~

20   ~~differences below the melting level. 99th percentile profiles show discrepancies as large as 20-25 dBZ in the mid-troposphere~~ ~~(Fig. 5b).~~ The Thompson scheme best reproduces the observed reflectivity vertical gradient, but absolute values of the reflectivity profiles aloft are closer between the simulations than between a single simulation and observed values. Additionally, there is no clear advantage in using bin versus bulk schemes for reproducing observed reflectivity profiles. ~~Although not shown, these 16Z differences are representative of the entire MCS lifecycle, and are consistent with numerous~~

25   ~~previous studies highlighting the high bias in simulated reflectivity for tropical deep convection.~~

5.2 MMD-T-*w*-TWC relationships

Isolating the role of potential simulated hydrometeor size biases in producing reflectivity biases requires controlling for *w* and TWC, which can also be biased and impact reflectivity. This is accomplished by examining relationships between MMD, *w*, T, and TWC.

30   **5.2.1 Observations**

Figure 6a shows average TWC (color-fill) as a function of T (ordinate) and *w* (abscissa) bins, while Fig. 6b and 6c show average MMD (color-fill) as a function of *w*-T and TWC-T bins, respectively. Note that the MMD color-fill scale is nonlinear.

Observed mean TWCs in Fig. 6a range from 1 to 3 g m^{-3} and generally increase with increasing _w for updrafts and increase with decreasing w for downdrafts_~~w~~. There is no clear relationship between TWC and T for T > -25 °C, although this is likely a result of biased observational sampling at warm temperatures. For T < -25 °C, TWC tends to increase for increasing ~~w~~T. Figure 6b shows a clear T-dependency for mean MMDs, which generally decrease with decreasing T. Observed mean MMDs range from 300-400 µm at T < -30 °C and approach ~1 mm at -10 °C ~~and w > 8 m s⁻¹~~. Most observed MMDs at T < -20 °C decrease slightly with increasing _w for updrafts_, but due to small sample sizes, no conclusions can be drawn for warmer T. MMD also appears to decrease slightly with downdraft velocity for a given T and are in fact higher for _w_ between -1 and 1 m s^{-1} compared to downdrafts, although downdraft samples are primarily limited to _w_ > -7 m s^{-1}. Figure 6c also shows that for T < -30 °C, MMDs generally decrease with increasing TWC, agreeing with L17~~6b~~, who show that observed MMDs in high IWC regions decrease with decreasing T and increasing TWC. However, this MMD-T-TWC relationship is sensitive to the type of MCS sampled. L17~~6b~~ discuss MMDs increasing with increasing TWC for Flights 12 and 13 (see Fig. 1) that sampled a long-lived, strongly cyclonic tropical low with persistent very deep convection at the center of the circulation. This inverse relationship is visible in Fig. 6c for T between -32 °C to -36 °C and between -24 °C and -28 °C. Interrogation of OAP images by L17~~6b~~ suggest that the dominant growth process for the larger MMDs during this event is vapor deposition, consistent with modest vertical wind speeds and little to no lightning during the event.

5.2.2 Differences between observations and simulations

Figure 7 shows relative differences between observed and simulated (simulated minus observed) mean TWC as a function of _w_-T bins (7a-c), as well as relative differences between observed and simulated mean MMD as a function of _w_-T (7d-f) and TWC-T (7g-i) bins. Differences between Thompson and observations are shown in the left column, between Morrison and observations in the middle column, and between FSBM and observations in the right column. Figure 7a-c shows that all simulations generally produce lesser than observed TWC for T < -24~~30~~ °C and ~~d~~ _w_ ~~<~~ between -5 and 7~~8~~ m s^{-1}, but generally larger TWC for warmer T, ~~or higher~~ _w_ > 7 m s^{-1}, ~~values~~and a few downdraft bins. The Thompson scheme reproduces observed mean TWCs with the greatest accuracy, generally remaining within 50% of those observed. Besides a few downdraft bins, ~~t~~The Morrison and FSBM schemes produce lesser TWC than observed by up to 50-100% at T < -36~~2~~ °C and _w_ < 5 m s^{-1}. However, for T > -30 °C or _w_ > 8 m s^{-1}, Morrison and FSBM produce slightly greater TWC than Thompson and observations. Recall that observed TWC uncertainty is ~0.1 g m^{-3} at -40 °C and ~0.3 g m^{-3} at -10 °C, however any potential bias over the entire range of samples is likely smaller and analyses of TWC differences between observations and simulations in this study focus on larger TWC values greater than 1 g m^{-3}.

Figure 7d-f~~Figure 8~~ shows relative differences between simulated and observed mean MMDs as a function of _w_ and T, where the simulated MMDs include all hydrometeor species together to mimic observations (Eq. 4). For T < -30 °C and _w_ < 8 m s^{-1}, the Thompson scheme ~~(Fig. 8a)~~ generally produces smaller than observed mean MMDs, whereas the Morrison and FSBM schemes produce larger than observed mean MMDs for the same _w_-T bins. Morrison produces the largest mean MMDs, commonly exceeding 100% larger than observed. For T > -30 °C, ~~or~~ _w_ > 7~~8~~ m s^{-1}, and _w_ < -1 m s^{-1}, all schemes produce

significantly larger mean MMDs than observed. All schemes exhibit a minimum in relative differences with observations for w between -1 m s^{-1} and 1 m s^{-1} and increasing relative differences with increasing absolute value of w.

~~Figure 9 is similar to Fig. 8~~Figure 7g-i also shows relative differences between observed and simulated MMDs, but~~,~~ ~~but~~ the abscissa is TWC rather than w. ~~MMDs as a function of T are similar to relationships in Fig. 8, by which T~~the Thompson scheme produces smaller than observed MMDs for T < -30 °C and TWC < 2 g m^{-3}. For T > -24 °C and TWC > 0.5 g m^{-3}, or for T between -24 °C and -40 °C and TWC > 2 g m^{-3}, Thompson produces larger than observed mean MMDs. The FSBM scheme ~~(Fig. 9c)~~ produces larger than observed mean MMDs across almost every TWC-T bin, but generally remains within 50% of observations for TWC < 1 g m^{-3}, while differences are exacerbated for larger TWC > 1 g m^{-3}. Morrison again produces the largest mean MMDs, exceeding 100% larger than observed across the majority of TWC-T bins.

Figures S2-S4 in the supplemental material show differences between observations and the three additional HIWC events (23 January, 2-3 February, and 7 February, see Fig. 1) simulated using the bulk schemes. Results from these additional simulations are very similar to those shown in Figure 7 for the 18 February event. This provides justification for using a single simulated event for comparison with the observations from all HIWC flights. Moreover, Figures S2-S4 show that there is little variability in MMD-w-T-TWC relationships as a function of simulated event for a given bulk microphysics scheme.

~~F~~Figure~~s~~ 7 ~~-9~~ suggest~~s~~ that high biases in simulated radar reflectivity are partly a result of too much condensate mass residing in particle diameters that are greater than observed MMDs of 0.4 to 1 mm rather than the lofting of excessive condensate mass and large particles by exaggerated vertical velocities alone. Proving that simulations have a hydrometeor size bias is difficult because of biased observational sampling that avoided the highest reflectivity convective cores. However, including sampling of these cores would still fail to bring observed and simulated MDs together for a given w or TWC since observed reflectivities aloft are significantly less than simulated, as was shown in Fig. 5. Additionally, analyzing minimum 90% MDs as a function of w or TWC can definitively establish a model hydrometeor size bias. The minimum 90% MD is defined as the minimum value among the distribution of 90% MDs within a given TWC-T or w-T bin. The minimum 90% MD is analyzed because it should not be impacted by the observational sampling bias (avoidance of lightning and reflectivities exceeding 40 dBZ) since it is associated with a lack of large, dense particles, whereas the observed MMDs shown in Figure 7 may be partly impacted by this observational sampling bias. If simulations fail to reproduce at least the minimum of the observed 90% MD distribution for a given bin, then there is greater confidence that a model bias exists. Figure 8~~10~~ shows joint histograms of simulated minus observed minimum 90% MDs as a function of TWC-T bins for bins where simulated values are greater than observed values. ~~Minimum 90% MDs are defined as the minimum value among all 90% MDs within a given TWC-T bin.~~ Simulations have a factor of 10^{2}~~3~~ more samples than observations, and therefore, bins with larger simulated than observed values almost certainly have 90% MDs that are biased high, implying again that too much mass is distributed to large particle sizes in simulations. Figure 8~~10~~a shows that minimum 90% MDs are larger in the Thompson scheme for TWC > 1 g m^{-3} and~~and~~ T between -2~~4~~0 °C and -40 °C or for TWC > 2 g m^{-3} and T between -8 °C and -20 °C. The Morrison scheme (Fig. 8~~10~~b) again produces the greatest differences with most bins for T between -20~~30~~ °C and -50 °C and

TWC > ~~1~~0.5 g m^{-3} exhibiting relative differences of 100% or greate~~rr~~. For T between -8 °C and -20 °C and TWC > 1 g m^{-3}, Morrison produces relative differences between 50% and 100 %. ~~.~~ FSBM (Fig. 8c)~~is closest to observations, but still~~ ~~contains~~ produces larger than observed minimum 90% MDs for T between -~~16~~30 °C and -50 °C and most TWC ~~> 1 g m^{-3}~~ bins, although the relative difference magnitudes are smaller relative to Thompson and Morrison.

5.3 MD-*w*-TWC relationships in specific temperature ranges

Figure 9~~11~~a-c shows average 10% MDs, MMDs, and 90% MDs, respectively, as functions of *w* for T between -32 °C and -40 °C. The Morrison and FSBM schemes produce larger than observed 10%, 50%, and 90% MDs for all *w*, with differences as large as 3 mm for the 90% MDs. The Thompson scheme is the only scheme that produces smaller than observed 10% MDs, whereas it captures observed MMDs remarkably well for *w* between 4 m s^{-1} and~~<~~ 10 m s^{-1}, although it diverges from observations outside of this range~~for *w* . > 10 m s^{-1}.~~ Although observed MMDs at larger *w* values may be biased low due to observational sampling that avoids the most intense convective cores, high biased Thompson reflectivities in this T range (see Fig. 5) suggest more representative observational sampling would not erase the MMD difference. This is even more evident for Thompson 90% MDs in Fig. 9~~11~~c, which reach values as high as 8 mm for 20 m s^{-1} updrafts. Similar~~ly~~ to Thompson, Morrison increases MMDs and 90% MDs with *w* for updrafts, but limits 90% MDs to less than 4 mm. FSBM produces an entirely different 90% MD-*w* relationship than the bulk schemes, where 90% MDs increase with increasing *w* ~~up to~~between 1 m s^{-1} and 6 m s^{-1}, but decrease with increasing *w* above this threshold. Similar relationships exist between MDs and TWC as for MDs and *w* in Fig. 9~~11~~, a result of TWC generally increasing with increasing *w* in this T range (see Fig. 6a and 1~~1~~3a-c in Sect. 5.4).

Figure 10~~2~~ shows the same information as Fig. 9~~11~~, but for a T range between -8 °C and -16 °C. Observations in Fig. 1~~2~~0 are plotted as individual data points (black diamonds) because of ~~few measurements at T > -20 °C~~ limited sampling of intense updrafts/downdrafts for T > -20 °C during the Darwin campaign. Limited measurements prevent conclusions about observed MD-*w* relationships in this T range, however some inferences can be drawn. ~~Note that m~~Much larger hydrometeor sizes are present than at colder temperatures, and thus the ordinate in Fig. 10~~2~~ is larger than in Fig. 9~~11~~ for MMDs and 90% MDs. Observed 10% MDs generally decrease with increasing *w* for updrafts, a relationship that all schemes capture to varying degrees. The Thompson scheme reproduces the observed 10% MD-*w* relationship best for *w* between~~<~~ 0 m s^{-1} and 4 m s^{-1}, but produces smaller than observed 10% MDs for higher *w*, related to excessive cloud droplet production. Morrison also produces smaller than observed 10% MDs for *w* > 4 m s^{-1}, but larger 10% MDs for smaller *w*. FSBM has far fewer cloud droplets (see Section 5.4) and produces larger than observed 10% MDs in updrafts and downdrafts~~across all *w*~~. Thompson and Morrison MMDs increase with increasing *w* and are larger than observed outside of quiescent regions, although Thompson produces much greater MMDs compared to Morrison for *w* >10 m s^{-1}. Thompson 90% MDs increase with increasing *w* for updrafts while FSBM 90% MDs decrease with *w* for *w* > 2 m s^{-1}. Morrison 90% MDs increase with increasing *w* ~~up to~~between 0 m s^{-1} and 6 m s^{-1}, but plateau at 6 mm for *w* > 6 m s^{-1}. While observed sample sizes in this temperature range are too small to confidently establish a 90% MD-*w* relationship, most observed 90% MDs are ~~much~~ smaller than simulated and 90% MDs

appear to increase with increasing w ~~in updrafts~~. Observed MDs as a function of TWC do not show a significant relationship in this temperature range, and thus MD-TWC relationships are not discussed here.

5.4 Model hydrometeor species partitioning

Recall that simulated MDs are computed from a composite MSD that incorporates all hydrometeor species in each microphysics scheme (Eq. 4). Therefore, differences between simulations and observations shown in Fig. 7-10 ~~8-12~~ may be partially explained by how mass is partitioned among species with different properties and how particle size is distributed by mass (MSD) for an individual hydrometeor species. In bulk schemes, these distributions are strongly modulated by the PSD parameters (N_0, μ, and λ) and the number of prognostic PSD moments (Varble et al., 2011). Although fewer assumptions are made regarding PSDs in the FSBM scheme, separate hydrometeor species with parameterized particle properties and microphysical processes still contribute to potential model biases. This section examines how MSD parameters (Eq. 1), assumed particle properties, and microphysical processes within each scheme may affect this partitioning of mass and particle size.

Figure 11 ~~3~~ shows w-T joint histograms for the 18 February 2014 simulated MCS. The color-fill in panels (a)-(c) is average TWC for Thompson, Morrison, and FSBM, respectively, and the color-fill in panels (d)-(f) is the average combined-hydrometeor MMD. Figures 11 ~~3~~a-c show that all schemes produce increasing TWC with increasing T and increasing downdraft or updraft velocity ~~w~~. The Morrison scheme produces the largest TWC with values of 4 g m^{-3} or greater for $w > 15$ m s^{-1} and T > -30 °C. Each scheme produces substantially different distributions of MMD as a function of T and w (Fig. 11 ~~3~~d-f), which is revealing when examined in conjunction with bulk mass. Both bulk schemes increase MMD with increasing T and increasing downdraft or updraft velocity ~~w and T~~. FSBM increases MMD with increasing T and decreasing updraft strength ~~w~~ for many bins, a relationship also exhibited by observations (Fig. 6b ~~e~~). Thompson produces the largest combined-hydrometeor MMDs, with maximum values approaching 1 cm for T > -10 °C and $w > 20$ m s^{-1}. Despite this, it shows a sharp T-dependency with mean MMDs smaller than 400 µm at T < -40 °C across all w values. Morrison average MMDs range from 1-3 mm for the majority of w-T bins, producing the largest sizes of any scheme for T < -30 °C, a feature also shown in Fig. 11 ~~9~~b. FSBM clearly produces the smallest MMDs on average, reaching a maximum of ~ 2 mm for weak w values and warm T. ~~between -12 °C and -16 °C.~~

Figure 14 ~~2~~ shows joint histograms organized similarly to Fig. 11 ~~3~~, but the color-fill in panels (a)-(c) is average snow (combined with cloud ice) water content (SWC), and the color-fill in panels (d)-(f) is average snow MMDs. The FSBM scheme produces the largest SWC with values of 1.5-2 g m^{-3} for T between -20 °C and -40 °C and $w > 10$ m s^{-1}. Morrison produces the lowest SWC, remaining below 1 g m^{-3} for all w-T bins, whereas Thompson produces values of 1.5-1.75 g m^{-3} for temperatures between -20 °C and -40 °C. The T-dependency in the Thompson snow MMDs is clear with a maximum of 1-3 mm for T > -20 °C and $w < 10$ m s^{-1}, decreasing to smaller than 0.5 mm at most T < -30 °C across all w. This pattern results from the T-dependent Thompson snow PSD parameterization coupled with the m-D relationship in Table 1 that forces a small, dense particle mode at colder T. Thompson's large SWC for T < -40 °C in Fig. 12 ~~4~~a largely controls the combined-hydrometeor

MMD at these temperatures (Fig. 13d) and is the cause for smaller than observed MMDs shown in Fig. 7~~8~~ and smaller than observed 10% MDs for cold temperatures in Fig. 9~~11~~. The FSBM snow MMD relationship with T and w is somewhat similar to Thompson since it also uses a variable snow density that accounts for high-density small crystals and low-density large aggregates. However, FSBM snow particles are not diagnosed to be small at cold T like they are in Thompson, and it produces MMDs up to 0.5 cm just above the melting level. Morrison produces the largest mean snow MMDs of all schemes across most w-T bins, with sizes approaching 1 cm for T > -10 °C and w > 5 m s^{-1}.

Figure 15~~3~~ is similar to Fig. 12~~4~~, but shows graupel water content (GWC) and graupel MMDs. All schemes increase graupel MMDs and GWC with increasing T and increasing downdraft or updraft velocity~~w and T~~. Thompson and FSBM GWC are distributed similarly, although FSBM generally produces more GWC than Thompson for a given w-T bin. Morrison produces the largest GWCs, which account for most of the Morrison TWCs shown in Fig. 11~~3~~b, strongly modulating Morrison combined-hydrometeor MMDs in Fig. 11~~3~~e. Thompson produces the largest graupel sizes of all the schemes, with many average graupel MMDs > 1 cm for high w and warm T. These large graupel sizes are a result of the diagnostic inverse relationship between graupel mass and N_0, forcing graupel to larger sizes as GWC increases (see Table 1). The Thompson scheme also shifts the fall-speed relationship to be more representative of hail at larger graupel sizes, resulting in faster sedimentation and limited GWC at cold T or small w (Fig. 13~~5~~a). However, this also creates very large graupel particles in intense updrafts in which graupel continues to be carried upward, which biases the combined-hydrometeor MMD, as was shown in Fig. 7~~8~~, 9~~11~~, and 10~~12~~. Morrison produces smaller graupel MMDs than snow MMDs (cf., Fig. 12~~4~~e and 13~~5~~e), where graupel sizes generally remain around 1-3 mm for most w-T bins. However, smaller graupel MMDs in conjunction with a slower terminal fall-speed results in less graupel sedimentation in the Morrison scheme and thus higher GWC than in the Thompson scheme. The Morrison combined-hydrometeor MMD (Fig. 11~~3~~e) is largely controlled by graupel because of the large GWCs, despite snow being the largest precipitating ice species because of low SWCs. FSBM produces the smallest mean graupel MMDs of all the schemes, generally remaining below 1 mm for most w-T bins and reaching sizes of ~2 mm just above the melting level. FSBM graupel MMDs are even smaller than FSBM snow MMDs (cf., Fig. 12~~4~~f and 13~~5~~f) and smaller than bulk scheme graupel MMDs. The unique FSBM feature of decreasing MMD with positive w is caused by the graupel MMDs being smaller than snow MMDs coupled with decreasing SWC and increasing GWC with increasing updraft strength~~w~~.

Average LWC (combining cloud water and rain) and liquid MMD joint histograms are shown in Fig. 14~~6~~. The bulk microphysics schemes distribute LWC similarly to one another as a function of T and w, although Thompson produces slightly more LWC than Morrison, especially just above the melting level. The FSBM scheme produces the largest LWCs for T between 0 °C and -8 °C, but much less supercooled LWC compared to bulk schemes for T < -8 °C. Thompson and Morrison liquid MMDs are relatively similar to one another, with cloud droplets dominating at T < -4 °C and raindrops dominating for T > -4 °C. Recall that for both bulk schemes, rain is a double-moment species and cloud water is a single-moment species. The FSBM scheme produces far different liquid MMDs as a function of w and T compared to the bulk schemes. For T > -4 °C, FSBM produces the smallest mean liquid MMDs and largest LWCs. For T < -4 °C, FSBM produces liquid MMDs more typical of small drizzle drops than cloud droplets. These unique features in the FSBM scheme may be the result of the explicit

30

activation of CCN and the maintenance of liquid supersaturation in FSBM that are absent in the bulk schemes. Aerosol consumption below the melting level in FSBM may cause faster and more efficient collision-coalescence than in the bulk schemes, allowing large raindrops to sediment more easily before reaching sub-freezing T. The smaller raindrop MMDs in FSBM may also be partly responsible for the smaller FSBM mean graupel MMDs compared to the bulk schemes (see Fig. 13f). The majority of graupel at T > -8 °C Figures 14c and 14f show that raindrop-sized MMDs control the large LWCs between 0 °C and -4 °C in updrafts, and IWC is negligible in updrafts at these temperatures. By -8 °C, nearly all liquid in updrafts is gone while GWC has increased dramatically, much more so than SWC (cf. Fig. 13f and Fig. 14f). Furthermore, GWC decreases with decreasing T below -8 °C, even though FSBM graupel terminal fall speeds (< 4 m s^{-1}) are weaker than the updraft speeds so that the graupel particles are carried upward. This suggests that most graupel production results from the freezing of raindrops, likely heterogeneously through interactions with pre-existing ice particles. is formed by freezing raindrops, and thus smaller raindrop MMDs are partly the cause for the smaller mean graupel MMDs produced in the FSBM scheme compared to the bulk schemes (Fig. 15f). Moreover, allowance of liquid supersaturation limits condensation in FSBM convective updrafts causing less LWC at T < -8 °C than in bulk schemes that condense all liquid supersaturation through saturation adjustment while applying a constant cloud droplet number concentration. Although results are not presented here for brevity, the aerosol aware Thompson scheme (Thompson and Eidhammer, 2014) with explicit CCN activation had smaller raindrops than the non-aerosol aware Thompson scheme for -4 °C > T > 0 °C with supercooled drizzle drops rather than cloud droplets as in FSBM, supporting the hypothesized reasons for differences between FSBM and the bulk schemes.

   Figures 113-164 show that condensate mass is partitioned differently among hydrometeor species for each scheme, which strongly impacts MD differences between schemes. MD differences also clearly rely on differences in the number of prognostic PSD moments for bulk schemes, mass-size relationships, PSD functions, fall-speed relationships, and microphysical process parameterizations. In particular, many studies have shown that hydrometeor sedimentation rates impact cloud and precipitation structure (e.g. Rutledge and Houze, 1987; Fovell and Ogurua, 1988; McCumber et al., 1991; Ferrier et al., 1995; Lynn et al., 2007; Thompson et al., 2008), and thus the assumed terminal velocity-size relationships described in Tables 1-3 may affect the model size biases presented. However, a sensitivity study of how variations in terminal velocity-size relationships impact MDs is beyond the scope of this study. Lastly, iIt is important to note that both graupel and snow MDs are larger than observed (cf. Fig. 6, 124, and 135), and therefore, overproduction of graupel in simulations is not solely responsible for the model hydrometeor size bias.

5.5 Connecting hydrometeor size biases to radar reflectivity biases

Equivalent Rayleigh reflectivity factor (Z_e) size distributions (ZSDs) describe how reflectivity is distributed by particle diameter. Simulated ZSDs are calculated given by the integrands of Eq. (6) and Eq. (7) for ice and liquid, respectively, such that the combined-hydrometeor ZSD ($Z_e(D)_{tot}$), as with combined-hydrometeor MSDs (Eq. 4), are the combined-hydrometeor Z_e distributions:

$$Z_e(D)_{tot} = 10^{18}\left[0.224\sum_{i=1}^{n}\left(\frac{6\alpha_i}{\pi\rho_w}\right)^2 D^{2\beta_i}N_i(D) + \sum_{j=1}^{m}D^6 N_j(D)\right] \qquad (8)$$

where n is the number of ice species and m is the number of liquid species. Observed ZSDs for ice particles are calculated using the integrand of Eq. (6) where the diameter definition is the area-equivalent diameter (D_{eq}). by multiplying the 6ᵗʰ power of the melted equivalent diameter (D_{eq}) by the PSD. For ice particles, D_{eq} is given by Eq. (5):

$$D_{eq} = \left[\frac{6\alpha}{\pi\rho_w}\right]^{\frac{1}{3}} D^{\frac{\beta}{3}} \qquad (5)$$

where ρ_w is the bulk density of water and $D_{eq} = D$ for liquid water particles. Observed ZSDs ($Z_e(D)$) for ice particles are then

5   given by Eq. (6):

$$Z_e(D) = 0.224 \times 10^{18} D_{eq}^6 N(D) \qquad (6)$$

where 10^{18} is a conversion factor from m⁶ to mm⁶ and 0.224 is a factor accounting for the different dielectric constants of ice and liquid, following Smith (1984). For liquid water species, Eq. (6) reduces to Eq. (7):

$$Z_e(D) = 10^{18} D^6 N(D) \qquad (7)$$

Simulated ZSDs, as with combined-hydrometeor MSDs (Eq. 4), are the combined-hydrometeor Z_e distributions:

$$Z_e(D)_{tot} = 10^{18}\left[0.224\sum_{i=1}^{n}\left(\frac{6\alpha_i}{\pi\rho_w}\right)^2 D^{2\beta_i}N_i(D) + \sum_{j=1}^{m}D^6 N_j(D)\right] \qquad (8)$$

where n is the number of ice species and m is the number of liquid species. Rayleigh reflectivity for each scheme is

10 then calculated by integrating the combined-hydrometeor ZSD ($Z_e(D)_{tot}$).

One caveat to interpreting differences between simulated and observed MSDs and ZSDs is that the observed MSDs for each 5-second flight sample are characterized by a single retrieved mass-diameter relationship power law, which may not work well in all situations, for example when relatively high density small and large particles co-exist with low density medium-sized particles. However, retrieval bias is likely limited by comparison of composite distributions, which are computed by taking

15 the mean mass for the composite MSD and mean equivalent reflectivity factor for the composite ZSD in each particle diameter bin. Furthermore, observed particles counts are more uncertain in the distribution tails at larger diameters (i.e. diameters > 3 mm). There is greater confidence placed in observed particle count, mass, and density for diameters smaller than 3 mm, by which an order of magnitude difference between observed and simulated masses or reflectivities at a given particle diameter requires greater than a 300% particle density error for the ZSD and 1000% particle density error for the MSD.

20 Figure 175 shows observed and simulated composite PSDs, MSDs, and ZSDs for T between -32 °C and -40 °C. All observed and simulated data points where TWC is between 2 and 2.5 g m⁻³ are included so that each composite distribution has approximately the same bulk mass. Simulated TWC constraints are limited to the observed diameter spectrum (maximum of 12.85 mm) for consistency among integrated distributions. No w constraint is used to increase the sample size of observed distributions. However, implementing $w > 1$ m s⁻¹ or $w > 5$ m s⁻¹ constraints increases the relative mass contained in larger

~~particle sizes in both simulations and observations, despite little change in MMD in observations (not shown).~~ Note that observed PSDs and MSDs include both D_{eq} (black) and D_{max} (grey) particle size definitions. All simulations struggle to reproduce the observed PSD using both the D_{eq} and D_{max} definitions with less particles than observed at diameters smaller than 0.5 mm. The Thompson scheme reproduces the observed profile reasonably well for diameters larger than 0.5 mm, while FSBM produces an order of magnitude more large particles than observed for diameters larger than 3 mm. Morrison produces the wrong PSD slope, distributing too many particles at sizes between 1 and 5 mm, but less particles than observed outside this range.

Composite MSDs (Fig. 1~~5~~7b) include symbols indicating MMDs (asterisks) and 90% MDs (triangles). Observations using the D_{eq} and D_{max} size definitions are relatively similar for sizes up to ~ 2.5 mm, but less mass is distributed at larger D_{max} values than at larger D_{eq} values. The observed MSD shows a prominent particle mass mode at ~ 300 µm, and remarkably, 90% of the > 2 g m^{-3} condensate mass observed is typically contained in particles with diameters < 1 mm. Thompson is the only simulation able to reproduce the prominent mass mode, whereas the FSBM and Morrison schemes shift it to larger sizes. Thompson reproduces observed mass for sizes up to 1.5 mm, but diverges significantly from observations at larger sizes, with differences as large as an order of magnitude for sizes between 2 and 5 mm. This is largely a result of the large graupel sizes produced by this scheme (see Fig. 1~~5~~3). FSBM captures the shape of the observed MSD reasonably well, but distributes too much mass at particles larger than ~ 500 µm, and too little mass at smaller diameters. The Morrison scheme again produces the wrong distribution slope, producing the largest discrepancies with observations for sizes between 1 and 4 mm. Simulated MMDs are closer to observed 90% MDs than observed MMDs, whereas simulated 90% MDs are several millimeters larger than observed. As noted in Sect. 5.4, this bias is not eliminated by considering simulated snow alone because both snow and graupel MDs are larger than observed MDs that consider all ice types.

Composite ZSDs for the D_{eq} definition alone are shown in Fig. 1~~7~~5c. The observed composite ZSD produces a reflectivity peak mode between 0.5 and 1 mm and a local minimum between 2 and 4 mm with a secondary mode at larger diameters. Both bulk schemes fail to capture the local minimum around 2-4 mm and produce larger reflectivities than retrieved between 2 and 5 mm where differences exceed an order of magnitude. FSBM also produces larger Z_e values compared to observations, despite being the only scheme that captures the local ZSD minimum. Notably, Thompson reproduces the observed ZSD up to 1.5 mm similarly to the MSD, but produces much higher reflectivities at larger diameters despite having only 10% of condensate mass distributed in these larger diameters. Recall that particle density errors of 1-2 orders of magnitude would be required to produce these large differences between observed and simulated Z_e, and thus there is confidence that these discrepancies are not entirely attributable to sampling bias, particularly for sizes smaller than 3 mm. It is also worth noting that comparison of median MSDs and ZSDs (not shown) shows that a significant fraction of observed distributions have little to no particles larger than a few millimeters, while this is rarely, if ever, produced in simulations for TWC > 2 g m^{-3}.

Composite distributions were additionally examined for a warmer temperature range (-8 °C to -20 °C), and differences between observations and simulations are still present and significant, although observational sample sizes are much smaller

at these temperatures. Notably, no scheme in the warmer T range is able to capture the prominent mass peak at ~ 300 µm that persists in observations, even at these warmer temperatures. This includes the Thompson scheme, suggesting that Thompson's ability to reproduce this mass mode at colder T, where SWC is the dominant bulk mass species, is due to the snow PSD parameterization that diagnostically forces particles to smaller sizes at colder T rather than more realistic physical process parameterizations. Overall, analysis of these ZSDs show that even though simulations often only distribute 10% of condensate mass at diameters larger than 2-3 mm, that small fraction of mass is often much greater than observed and greatly biases radar reflectivity.

6 Conclusions

Properly representing cloud microphysical processes and hydrometeor properties in microphysics parameterizations is vital to improving simulations of clouds and precipitation, but identifying sources of model bias is difficult given the complexity of nonlinear interactions between dynamics and microphysics within the model. This study differs from previous studies by controlling for vertical velocity, bulk condensate mass, and temperature to isolate the contribution of hydrometeor size from the contribution of excessive condensate mass in overly strong or large convective updrafts to the well-known high bias in simulated tropical deep convective radar reflectivity.

Data collected during the first HAIC-HIWC field campaign held in Darwin, Australia, in 2014, are compared with three WRF simulations of a mesoscale convective system that passed through Darwin on 18 February. The simulations vary only by the microphysics scheme employed (Thompson, Morrison, and FSBM). While a reflectivity bias in simulated tropical convection has been previously shown to exist in bulk schemes, this study shows that the bias exists in a bin scheme (FSBM) as well. Simulated MMDs and upper percentile MDs are larger than observed in every microphysics scheme for a given w, TWC, and T condition. Many TWC-T and w-T conditions exist in each scheme where the minimum 90% MD is larger than observed, despite simulated samples sizes being ~ 10^{23} times larger than observations.

For temperatures between -30 °C and -50 °C, the Thompson scheme best reproduces observed MMDs for a given TWC or w. Vapor-grown ice particles largely control the bulk mass at these temperatures, and the Thompson scheme uses a unique snow PSD parameterization combined with a m-D relationship that forces snow mass into smaller sizes with decreasing T. Although these small, dense snow particles are diagnosed rather than produced by a microphysical process, they suggest that using a m-D relationship in which vapor-grown ice particle density decreases with particle size combined with an appropriate PSD function can nearly reproduce observations at cold temperatures where vapor-grown ice contributes most to condensate mass. However, the Thompson scheme produces larger than observed MMDs for temperatures between -10 °C and -30 °C, especially for high TWC or w, where graupel largely controls condensate mass. The graupel size is diagnostically increased as graupel bulk mass increases, pushing graupel to large sizes. While large graupel particles have a hail-like fall-speed that sediments them out of updrafts more quickly than in the Morrison and FSBM schemes, their sizes become so large that only a small graupel bulk mass easily high biases radar reflectivity.

The Morrison scheme allows for much greater variability in graupel and snow size since it predicts N for both species, making it a two-moment ice scheme. This shifts graupel to smaller mean sizes and snow to larger mean sizes than in the Thompson scheme. Indeed, snow MMDs are on average larger than graupel MMDs in the Morrison scheme, and both are larger than observed MMDs for most T-w-TWC conditions. However, fall--speeds of Morrison graupel that are significantly slower than Thompson graupel fall speeds result in much of the bulk mass being controlled by GWC in updraft cores, even at temperatures down to -50 °C. This causes the combined-hydrometeor MMDs in Morrison to be largely controlled by graupel.

 The FSBM scheme has a fundamental advantage over the bulk schemes in that it does not assume a PSD shape and computes microphysical process rates separately for different hydrometeor size bins. Similarly to the Morrison scheme, snow MMDs are larger than graupel MMDs in the FSBM scheme, which combined with large amounts of SWC, indicate that snow is largely responsible for size biases in FSBM. Notably, graupel sizes in FSBM generally remain below 1 mm for most T-w-TWC bins. These smaller graupel sizes may result from two processes unique to the FSBM scheme: explicit CCN nucleation and maintenance of liquid supersaturation. These processes may aid in reducing the size of lofted raindrops that freeze upon collision with ice particles to form most graupel, while less supercooled liquid limits additional riming.

Perhaps most revealing are features from composite mass and equivalent Rayleigh reflectivity factor distributions as a function of diameter. A prominent mass mode at ~ 300 µm exists in observed MSDs regardless of T, w, or TWC constraints, but rarely are any of the schemes able to reproduce this feature, with the exception of the Thompson scheme at temperatures colder than -30 °C. Otherwise, all schemes produce too much mass at large particle diameters, even for TWC < 1 g m^{-3}, although discrepancies with observations are enhanced for larger TWCs and higher w values. The excess simulated mass at diameters > 1 mm leads to reflectivity factors that are higher than observed by up to two orders of magnitude in some diameter ranges.

Ultimately, all simulations fail to reproduce observed hydrometeor size distributions in which the majority of bulk mass is distributed at sub-mm sizes like commonly observed. Bulk scheme mass distributions are sensitive to assumed hydrometeor properties including the PSD function and the mass-size relationship. The bin scheme failures show that additional causes of hydrometeor size biases are likely related to species partitioning and parameterization of microphysical processes. Biases resulting from microphysical processes are likely present in bulk schemes as well, but further research is needed to determine how much of the bias results from microphysical process parameterization errors versus diagnosed single particle and PSD properties. Future work using data from the second HAIC-HIWC phase held in Cayenne, French Guiana, in 2015 will enable more in-depth evaluation of biases in deep convection at temperatures warmer than -15 °C, where size biases appear to originate in convective updrafts.

References

Ackerman, A. S., Fridlind, A. M., Grandin, A., Dezitter, F., Weber, M., Strapp, J. W., and Korolev, A. V.: High ice water content at low radar reflectivity near deep convection – Part 2: Evaluation of microphysical pathways in updraft parcel simulations, Atmos. Chem. Phys., 15, 11729-11751, doi: 10.5194/acp-15-11729-2015, 2015.

Baumgardner, D., Brenguier, J. L., Bucholtz, A., Coe, H., DeMott, P., Garrett, T. J., Gayet, J. F., Hermann, M., Heymsfield, A., Korolev, A., Krämer, M., Petzold, A., Strapp, J. W., Pilewskie, P., Taylor, J., Twohy, C., Wendisch, M., Bachalo, W., and Chuang, P.: Airborne instruments to measure atmospheric aerosol particles, clouds and radiation: A cook's tour of mature and emerging technology, Atmos. Res., 102, 10-29, doi: 10.1016/j.atmosres.2011.06.021, 2011.

Bechtold, P., Redelsperger, J.-L., Beau, I., Blackburn, M., Brinkop, S., Grandper, J.-Y., Grant, A., Gregory, D., Guichard, F., How, C., and Ioannidou, E.: A GCSS model intercomparison for a tropical squall line observed during TOGA-COARE. II: Intercomparison of single-column models and a cloud-resolving model, Q. J. R. Meteorolog. Soc., 126, 865-888, doi: 10.1002/qj.49712656405, 2000.

Blahak, U.: RADAR_MIE_LM and RADAR_MIELIB— Calculation of radar reflectivity from model output. Internal Rep., Institute for Meteorology and Climate Research, University/Research Center Karlsruhe, 150 pp., 2007.

Blossey, P. N., Bretherton, C. S., Cetrone, J., and Kharoutdinov, M.: Cloud-resolving model simulations of KWAJEX: Model sensitivities and comparisons with satellite radar observations, J. Atmos. Sci., 64, 1488-1508, doi: 10.1175/JAS3982.1, 2007.

Caine, S., Lane, T. P., May, P. T., Jakob, C., Siems, S. T., Manton, M. J., and Pinto, J.: Statistical assessment of tropical convection-permitting model simulations using a cell-tracking algorithm, Mon. Weather Rev., 141, 557-581, doi: 10.1175/MWR-D-11-00274.1, 2013.

Chen, S. and Cotton, W.: The sensitivity of a simulated extratropical mesoscale convective system to longwave radiation and ice-phase microphysics, J. Atmos. Sci., 45, 3897–3910, doi: 10.1175/1520-0469(1988)045<3897:TSOASE>2.0.CO;2, 1988.

Chen, F. and Dudhia, J.: Coupling an advanced land-surface/ hydrology model with the Penn State/ NCAR MM5 modeling system. Part I: Model description and implementation, Mon. Weather Rev., 129, 569–585, doi: 10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2, 2001.

Cox, G. P.: Modelling precipitation in frontal rainbands, Q. J. R. Meteorolog. Soc., 114, 115–127, doi: 10.1002/qj.49711447906, 1998.

Davison, C. R., MacLeod, J., Strapp, J. W., and Buttsworth, D.: Isokinetic total water content probe in a naturally aspirating configuration: Initial aerodynamic design and testing, 46th AIAA Aerospace Sciences Meeting and Exhibit, Aerospace Sciences Meetings, Reno, Nevada, 7-10 January 2008, available at http://dx.doi.org/10.2514/6.2008-435, 2008.

Davison, C. R., Strapp, J. W., Lilie, L. E., Ratvasky, T. P., and Dumont, C.: Isokinetic TWC evaporator probe: Calculations and systemic error analysis, 8th AIAA Atmospheric and Space Environments Conference, Washington, DC, 13-17 June 2016, available at http://dx.doi.org/10.2514/6.2016-4060, 2016.

Dezitter, F., Grandin, A., Brenguier, J.-L., Hervy F., Schlager, H., Villedieu, P., and Zalamansky, G., 2013: HAIC – High Altitude Ice Crystals. 5th AIAA Atmospheric and Space Environments Conference, San Diego, CA, 24-27 June 2013, available at http://arc.aiaa.org/doi/abs/10.2514/6.2013-2674, 2013.

Dudhia, J.: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model, J. Atmos. Sci., 46, 3077–3107, doi: 10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2, 1989.

Ferrier, B. S., A double-moment multiple-phase four-class bulk ice scheme. Part I: Description, J. Atmos. Sci., 15, 249-280, doi: 10.1175/1520-0469(1994)051<0249:ADMMPF>2.0.CO;2, 1994.

Ferrier, B.S.,Tao, W., and Simpson, J.: A Double-Moment Multiple-Phase Four-Class Bulk Ice Scheme. Part II: Simulations of Convective Storms in Different Large-Scale Environments and Comparisons with other Bulk Parameterizations, J. Atmos. Sci., 52, 1001–1033, doi: 10.1175/1520-0469(1995)052<1001:ADMMPF>2.0.CO;2, 1995.

Field, P. R., Wood, R., Brown, P., Kaye, P., Hirst, E., Greenaway, R., and Smith, J.: Ice particle interarrival times measured with a fast FSSP, J. Atmos. Oceanic Technol., 20, 249–261, doi: 10.1175/1520-0426(2003)020<0249:IPITMW>2.0.CO;2, 2003.

Field, P. R., Hogan, R. J., Brown, P. R. A., Illingworth, A. J., Choularton, T. W., and Cotton, R. J.: Parametrization of ice-particle size distributions for mid-latitude stratiform cloud, Q. J. R. Meteorolog. Soc., 131, 1997–2017, doi: 10.1256/qj.04.134, 2005.

Field, P.R., Heymsfield, A. J., and Bansemer, A.: Shattering and particle interarrival times measured by optical array probes in ice clouds, J. Atmos. Oceanic Technol., 23, 1357-1371, doi: 10.1175/JTECH1922.1, 2006.

Fovell, R. G. and Ogura, Y.: Numerical simulation of a midlatitude squall line in two dimensions, J. Atmos. Sci.,45, 3846-3879, doi: 10.1175/1520-0469(1988)045<3846:NSOAMS>2.0.CO;2., 1988.

Franklin, C. N., Protat, A., Leroy, D., and Fontaine, E.: Controls on phase composition and ice water content in a convection-permitting model simulation of a tropical mesoscale convective system, Atmos. Chem. Phys., 16, 8767-8789, doi: 10.5194/acp-16-8767-2016, 2016.

Fridlind, A. M., Ackerman, A., Petch, J., Field, P., Hill, A., McFarquhar, G., Xie, S., and Zhang, M.: ARM/GCSS/SPARC TWP-ICE CRM intercomparison study, NASA Tech. Memo., TM-2010-215858, 2010.

Fridlind, A. M., Ackerman, A. S., Grandin, A., Dezitter, F., Weber, M., Strapp, J. W., Korolev, A. V., and Williams, C. R.: High ice water content at low radar reflectivity near deep convection – Part 1: Consistency of in situ and remote-sensing observations with stratiform rain column simulations, Atmos. Chem. Phys., 15, 11713-11728, doi: 10.5194/acp-15-11713-2015, 2015.

Garrett, T., Navarro, B., Twohy, C., Jensen, E., Baumgardner, D., Bui, P., Gerber, H., Herman, R., Heymsfield, A., Lawson, P., Minnis, P., Nguyen, L., Poellot, M., Pope, S., Valero, F., and Weinstock, E.: Evolution of a Florida cirrus anvil, J. Atmos. Sci., 62, 2352–2372, doi: 10.1175/JAS3495.1, 2005.

Gilmore, M., Straka, J., and Rasmussen, E.: Precipitation uncertainty due to variations in precipitation particle parameters within a simple microphysics scheme, Mon. Weather Rev., 132, 2610–2627, doi: 10.1175/MWR2810.1, 2004.

Hartmann, D., Ockert-Bell, M., and Michelsen, M.: The effect of cloud type on earth's energy balance: Global analysis, J. Clim., 5, 1281–1304, doi: 10.1175/1520-0442(1992)005<1281:TEOCTO>2.0.CO;2, 1992.

Heymsfield, A. J.: Properties of tropical and midlatitude ice cloud particle ensembles. Part I: Median mass diameters and terminal velocities, J. Atmos. Sci., 60, 2573-2591, doi: 10.1175/1520-0469(2003)060<2573:POTAMI>2.0.CO;2., 2003a.

Heymsfield, A. J.: Properties of Tropical and Midlatitude Ice Cloud Particle Ensembles. Part II: Applications for Mesoscale and Climate Models, J. Atmos. Sci., 60, 2592–2611, doi: 10.1175/1520-0469(2003)060<2592:POTAMI>2.0.CO;2, 2003b.

Heymsfield, A. J.: On measurements of small ice particles in clouds, Geophys. Res. Lett., 34, doi:10.1029/2007GL030951, 2007.

Heymsfield, A. J. and McFarquhar, G.: High albedos of cirrus in the tropical Pacific Warm Pool: Microphysical interpretations from CEPEX and from Kwajalein, Marshall Islands, J. Atmos. Sci., 53, 2424–2451, doi: 10.1175/1520-0469(1996)053<2424:HAOCIT>2.0.CO;2, 1996.

Heymsfield, A. and Willis, P.: Cloud Conditions Favoring Secondary Ice Particle Production in Tropical Maritime Convection, J. Atmos. Sci., 71, 4500–4526, doi: 10.1175/JAS-D-14-0093.1, 2014.

Heymsfield, A. J., Miloshevich, L., Schmitt, C., Bansemer, A., Twohy, C., Poellot, M., Fridlind, A., and Gerber, H.: Homogeneous ice nucleation in subtropical and tropical convection and its influence on cirrus anvil microphysics, J. Atmos. Sci., 62, 41–64, doi: 10.1175/JAS-3360.1, 2005.

Heymsfield, A. J., Bansemer, A., Heymsfield, G., and Fierro, A.: Microphysics of maritime tropical convective updrafts at temperatures from $-20°$ to -60, J. Atmos. Sci., 66, 3530–3562, doi: 10.1175/2009JAS3107.1, 2009.

Heymsfield, G., Tian, L., Heymsfield, A. J., Li, L., and Guimond, S.: Characteristics of deep tropical and subtropical convection from nadir-viewing high-altitude airborne Doppler radar, J. Atmos. Sci., 67, 285–308, doi: 10.1175/2009JAS3132.1, 2010.

Janjic, A. I.: The step-mountain Eta Coordinate Model: Further developments of the convection, viscous sublayer, and turbulence closure schemes, Mon. Weather Rev., 122, 927-945, doi: 10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2., 1994.

Jorgensen, D. P. and LeMone, M. A.: Vertical velocity characteristics of oceanic convection, J. Atmos. Sci., 46, 621-640, doi: 10.1175/1520-0493(1999)127<1056:VVCODC>2.0.CO;2, 1989.

Kain, J. S.: The Kain-Fritsch Convective Parameterization: An update, J. Appl. Meteorol., 43, 170-181, doi: 10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2, 2004.

Keenan, T. D., Glasson, K., Cummings, F., Bird, T. S., Keeler, J., and Lutz, J.: The BMRC/NCAR C-Band Polarimetric (C-POL) Radar System, J. Atmos. Oceanic Technol., 15, 871–886, doi: 10.1175/1520-0426(1998)015<0871:TBNCBP>2.0.CO;2, 1998.

Khain, A. P., Beheng, K. D., Heymsfield, A., Korolev, A., Krichak, S. O., Levin, Z., Pinsky, M., Phillips, V., Prabhakaran, T., Teller, A., van den Heever, S. C., and Yano, J.-I.: Representation of microphysical processes in cloud-resolving models: Spectral (bin) microphysics versus bulk parameterization, Rev. Geophys., 53, 247-322, doi: 10.1002/2014RG000468, 2015.

Korolev, A. and Sussman, B.: A technique for habit classification of cloud particles, J. Atmos. Oceanic Technol., 17, 1048–1057, doi: 10.1175/1520-0426(2000)017<1048:ATFHCO>2.0.CO;2, 2000.

Korolev, A. and Isaac, G. A.: Shattering during sampling by OAPs and HVPS. Part 1: Snow particles, J. Atmos. Oceanic Technol., 22, 528-542, doi: 10.1175/JTECH1720.1, 2005.

Korolev, A., and Field, P. R.: Assessment of the performance of the inter-arrival time algorithm to identify ice shattering artifacts in cloud particle probe measurements, Atmos. Meas. Tech., 8, 761-777, doi: 10.5194/amt-8-761-2015, 2015.

Lang, S., Tao, W.-K., Simpson, J., Cifelli, R., Rutledge, S., Olson, W., and Halverson, J.: Improving simulations of convective systems from TRMM LBA: Easterly and westerly regimes, J. Atmos. Sci., 64, 1141-1164, doi: 10.1175/JAS3879.1, 2007.

Lang., S., Tao, W.-K., Zeng, X., Li, Y.: Reducing the biases in simulated radar reflectivities from a bulk microphysics scheme: Tropical convective systems, J. Atmos. Sci., 68, 2306-2320, doi: 10.1175/JAS-D-10-05000.1, 2011.

Lang, S., Tao, W.-K., Chern, J., Wu, D., and Li, X.: Benefits of a fourth ice class in the simulated radar reflectivities of convective systems using a bulk microphysics scheme, J. Atmos. Sci., 71, 3583–3612, doi: 10.1175/JAS-D-13-0330.1, 2014.

Lawson, R. P.: Effects of ice particles shattering on the 2D-S probe, Atmos. Meas. Tech., 4, 1361-1381, doi:10.5194/amt-4-1361-2011, 2011.

Lawson, R. P., Angus, L. J., and Heymsfield, A. J.: Cloud Particle Measurements in Thunderstorm Anvils and Possible Weather Threat to Aviation, J. Aircr., 35, 113-121, doi: 10.2514/2.2268, 1998.

Lawson, R. P., O'Connor, D., Amarzly, P., Weaver, K., Baker, B., Mo., Q., and Jonsson, H.: The 2D-S (Stereo) Probe: Design and preliminary tests of a new airborne, high-speed, high-resolution particle imaging probe, J. Atmos. Oceanic Technol., 23, 1462-1477, doi: 10.1175/JTECH1927.1, 2006.

Lawson, R., Woods, S., and Morrison, H.: The microphysics of ice and precipitation development in tropical cumulus clouds, J. Atmos. Sci., 72, 2429–2445, doi: 10.1175/JAS-D-14-0274.1, 2015.

Leroy, D., Fontaine, E., Schwarzenboeck, A., and Strapp, J. W.: Ice crystal sizes in high ice water content clouds. Part I: On the computation of median mass diameter from in situ measurements, J. Atmos. Oceanic Technol., 33, 2461–2476, doi: 10.1175/JTECH-D-15-0151.1, 2016a.

Leroy, D., Fontaine, E., Schwarzenboeck, A., Strapp, J. W., Korolev, A., McFarquhar, G., Dupuy, R., Gourbeyre, C., Lilie, L., Protat, A., Delanoë, J., Dezitter, F., and Grandin, A.: Ice crystal sizes in high ice water content clouds. Part 2: Statistics of mass diameter percentiles in tropical convection observed during the HAIC/HIWC project, J. Atmos. Oceanic Technol., 34, 117-136in press, doi: 10.1175/JTECH-D-15-0246.1, 20176b.

Li, Y., Zipser, E. J., Krueger, S. K., and Zulauf, M. A.: Cloud-resolving modeling of deep convection during KWAJEX: Part 1: Comparison to TRMM satellite and ground-based observations, Mon. Weather Rev., 136, 2699-2712, doi: 10.1175/2007MWR2258.1, 2008.

Lynn, B. H., Khain, A. P., Dudhia, J., Rosenfeld, D., Pokrovsky, A., and Seifert, A.: Spectral (bin) microphysics coupled with a mesoscale model (MM5). Part I: Model description and first results, Mon. Weather Rev., 133, 44-58, doi: 10.1175/MWR-2840.1, 2005.

Lynn, B., Khain, A., Rosenfeld, D., and Woodley, W. L.: Effects of aerosols on precipitation from orographic clouds, J. Geophys. Res., 112, D10225, doi:10.1029/2006JD007537, 2007.

Martin, G., Johnson, D., and Spice, A.: The measurement and parameterization of effective radius of droplets in warm stratocumulus clouds, J. Atmos. Sci., 51, 1823–1842, doi: 10.1175/1520-0469(1994)051<1823:TMAPOE>2.0.CO;2, 1994.

Mason, J. G., Strapp, J. W., and Chow, P.: The ice particle threat to engines in flight, 44th AIAA Aerospace Sciences Meeting, Reno, Nevada, 9-12 January 2006, AIAA-2006-206, 2006.

Matsui, T., Zeng, X., Tao, W.-K., Masunaga, H., Olson, W. S., and Lang, S.: Evaluation of long-term cloud-resolving model simulations using satellite radiance observations and multifrequency satellite simulations, J. Atmos. Oceanic Technol., 26, 2185-2197, doi: 10.1175/2008JTECHA1168.1, 2009.

McCumber, M., Tao, W.-K., Simpson, J., Penc, R., and Soong, S.: Comparison of ice-phase microphysical parameterization schemes using numerical simulations of tropical convection, J. Appl. Meteorol., 30, 985–1004, doi: 10.1175/1520-0450-30.7.985, 1991.

McFarquhar, G. and Heymsfield, A.: Microphysical characteristics of three anvils sampled during the Central Equatorial Pacific Experiment, J. Atmos. Sci., 53, 2401–2423, doi: 10.1175/1520-0469(1996)053<2401:MCOTAS>2.0.CO;2, 1996.

McFarquhar, G. and Heymsfield, A.: Parameterization of tropical cirrus ice crystal size distributions and implications for radiative transfer: Results from CEPEX, J. Atmos. Sci., 54, 2187–2200, doi: 10.1175/1520-0469(1997)054<2187:POTCIC>2.0.CO;2, 1997.

McFarquhar, G., Zhang, H., Heymsfield, G., Halverson, J., Hood, R., Dudhia, J., and Marks, F.: Factors affecting the evolution of Hurricane Erin (2001) and the distributions of hydrometeors: Role of microphysical processes, J. Atmos. Sci., 63, 127–150, doi: 10.1175/JAS3590.1, 2006.

Milbrandt, J. A., and Yau, M. K.: A multimoment bulk microphysics parameterization. Part II: A proposed three-moment closure and scheme description, J. Atmos. Sci., 62, 3065-3081, doi: 10.1175/JAS3535.1, 2005.

Milbrandt, J. A., and Yau, M. K.: A multimoment bulk microphysics parameterization. Part III: Control simulation of a hailstorm, J. Atmos. Sci., 63, 3114-3136, doi: 10.1175/JAS3816.1, 2006.

Mitchell, D. L., Zhang, R., and Pitter, R. L.: Mass-dimensional relationships for ice particles and the influence of riming on snowfall rates, J. Appl. Meteorol., 29, 153-163, doi: 10.1175/1520-0450(1990)029<0153:MDRFIP>2.0.CO;2, 1990.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, J. Geophys. Res., 102, 16663–16682, doi: 10.1029/97JD00237, 1994.

Morrison, H. and Grabowski, W.: A novel approach for representing ice microphysics in models: description and tests using a kinematic framework, J. Atmos. Sci., 65, 1528–1548, doi: 10.1175/2007JAS2491.1, 2008.

Morrison, H., Thompson, G., and Tatarskii, V.: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes, Mon. Weather Rev., 137, 991-1007, doi: 10.1175/2008MWR2556.1, 2009.

Nesbitt, S. W., Cifelli, R., and Rutledge, S. A.: Storm morphology and rainfall characteristics of TRMM precipitation features, Mon. Weather Rev., 134, 2702-2721, doi: 10.1175/MWR3200.1, 2006.

Puri, K., Dietachmayer, G., Steinle, P., Dix, M., Rikus, L., Logan, L., Naughton, M. Tingwell, C., Xiao, Y., Barras, V., Bermous, I., Bowen, R., Deschamps, L., Franklin, C. N., Fraser, J., Glowacki, T. J., Harris, B., Lee, J., Le, T., Roff, G. L., Sulaiman, A., Sims, H., Sun, X., Sun, Z., Zhu, H., Chattopadhyay, M., and Engel, C.: Implementation of the initial ACCESS numerical weather prediction system, Aust. Meteorol. Oceanogr. J., 63, 265-284, 2013.

Reisner, R., Rasmussen, R. M., and Bruintjes, R. T.: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model, Q. J. R. Meteorolog. Soc., 124, 1071-1107, doi: 10.1002/qj.49712454804, 1998.

Rutledge, S.A. and Houze, R. A.: A Diagnostic Modelling Study of the Trailing Stratiform Region of a Midlatitude Squall Line, J. Atmos. Sci., 44, 2640–2656, doi: 10.1175/1520-0469(1987)044<2640:ADMSOT>2.0.CO;2, 1987.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 3, NCAR Tech. Note NCAR/TN-475+STR, 113 pp., doi: 10.5065/D68S4MVH, 2008.

Smith, P. L.: Equivalent radar reflectivity factors for snow and ice particles, J. Appl. Meteorol., 23, 1258-1260, doi: 10.1175/1520-0450(1984)023<1258:ERRFFS>2.0.CO;2, 1984.

Stith, J. L., Dye, J. E., Bansemer, A., Heymsfield, A. J., Grainger, C. A., Petersen, W. A., and Cifelli, R.: Microphysical observations of tropical clouds, J. Appl. Meteorol., 41, 97-117, doi: 10.1175/1520-0450(2002)041<0097:MOOTC>2.0.CO;2, 2002.

Stith J. L., Haggerty, J. A., and Heymsfield, A. J.: Microphysical characteristics of tropical updrafts in clean conditions, J. Appl. Meteorol., 43, 779-794, doi: 10.1175/2104.1, 2004.

Strapp, J. W., Isaac, G. A., Korolev, A., Ratvasky, T., Potts, R., May, P., Protat, A., Minnis, P., Ackerman, A., Fridlind, A., Haggerty, J., and Riley, J.: The High Ice Water Content (HIWC) study of deep convective clouds: Science and technical plan, FAA Rep. DOT/FAA/TC-14/31, 105 pp., available at http://www.tc.faa.gov/its/worldpac/techrpt/tc14-31.pdf, 2016.

Strapp, J. W., Lilie, L. E., Ratvasky, T. P., Davison, C., and Dumont, C.: Isokinetic TWC evaporator probe: Development of the IKP2 and performance testing for the HAIC-HIWC Darwin 2014 and Cayenne-2015 field campaigns, 8th AIAA Atmospheric and Space Environments Conference, Washington, DC, 13-17 June 2016, doi: 10.2514/6.2016-4059, 2016.

Tao, W.-K., and Moncrieff, M. W.: Multiscale cloud system modeling, Rev. Geophys., 47, RG4002, doi: 10.1029/2008RG000276, 2009.

Thompson, G., Field, P. R., Rasmussen, R. M., and Hall, W. D.: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization, Mon. Weather Rev., 136, 5095-5115, doi: 10.1175/2008MWR2387.1, 2008.

Thompson, G. and Eidhammer, T.: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone, J. Atmos. Sci., 71, 3636-3658, doi: 10.1175/JAS-D-13-0305.1, 2014.

Varble, A., Fridlind, A. M., Zipser, E. J., Ackerman, A. S., Chaboureau, J.-P., Fan, J., Hill, A., McFarlane, S. A., Pinty, J.-P., and Shipway, B.: Evaluation of cloud-resolving model intercomparison simulations using TWP-ICE observations: Precipitation and cloud structure, J. Geophys. Res., 116, D12206, doi:10.1029/2010JD015180, 2011.

Varble, A., Zipser, E. J., Fridlind, A. M., Zhu, P., Ackerman, A. S., Chaboureau, J.-P., Collis, S., Fan, J., Hill, A., and Shipway, B., Evaluation of cloud-resolving and limited area model intercomparison simulations using TWP-ICE observations: 1. Deep convective updraft properties, J. Geophys. Res. Atmos., 119, 13891-13918, doi:10.1002/2013JD021371, 2014a.

Varble, A., Zipser, E. J., Fridlind, A. M., Zhu, P., Ackerman, A. S., Chaboureau, J.-P., Fan, J., Hill, A., Shipway, B., and Williams, C.: Evaluation of cloud-resolving and limited area model intercomparison simulations using TWP-ICE observations: 2. Precipitation Microphysics, J. Geophys. Res. Atmos., 119, 13919-13945, doi: 10.1002/2013JD021372, 2014b.

Wang, W. and Liu, X.: Evaluating deep updraft formulation in NCAR CAM3 with high-resolution WRF simulations during ARM TWP-ICE, Geophys. Res. Lett., 36, L04701, doi: 10.1029/2008GL036692, 2009.

Wang, Y., Long, C., Leung, L., Dudhia, J., McFarlane, S., Mather, J., Ghan, S., and Liu, X.: Evaluating regional cloud-permitting simulations of the WRF model for the Tropical Warm Pool International Cloud Experiment (TWP-ICE), Darwin, 2006, J. Geophys. Res., 114, D21203, doi: 10.1029/2009JD012729, 2009.

Westbrook, C. D., Ball, R. C., Field, P. R., and Heymsfield, A. J.: Universality in snowflake aggregation, Geophys. Res. Lett., 31, L15104, doi:10.1029/2004GL020363, 2004.

Yang, J., Wang, Z., Heymsfield, A. J., and Luo, T.: Liquid–ice mass partition in tropical maritime convective clouds, J. Atmos. Sci., 73, 4959–4978, doi: 10.1175/JAS-D-15-0145.1, 2016.

Zhu, P., Dudhia, J., Field, P. R., Wapler, K., Fridlind, A., Varble, A., Zipser, E. J., Petch, J., Chen, M., and Zhu, Z.: A limited area model (LAM) intercomparison study of a TWP-ICE active monsoon mesoscale convective event, J. Geophys. Res., 117, D11208, doi:10.1029/2011JD016447, 2012.

Zipser, E. J., and Lutz, K. R.: The vertical profile of radar reflectivity of convective cells: A strong indicator of storm intensity and lightning probability?, Mon. Weather Rev., 122, 1751-1759, doi: 10.1175/1520-0493(1994)122<1751:TVPORR>2.0.CO;2, 1994.

Table 1. Thompson mass-size relationship ($m = \alpha D^\beta$), terminal velocity-size relationship ($v =[\rho_0/\rho]^{1/2}cD^d exp[-fD])$, and gamma PSD parameter values used for each species. N_c used for cloud water μ and N_0 calculations is a constant 100 cm^{-3}. N_0 equations for double-moment species are determined by prognostic N and q. The bimodal gamma distribution used for the Thompson snow PSD may be found in Thompson et al. (2008), Eq. (1). The f parameter in the terminal velocity-size relationship follows Ferrier (1994).

								Thompson m-D, v-D, and PSD ~~MSD~~ parameters	
Species	Prognostic Variables	ρ [kg m^{-3}]	α	β	c	d	f	N_0 [m^{-4}]	μ
Snow	q_s	$\frac{6\alpha_s}{\pi}D^{\beta_s-3}$	0.069	2	40	0.55	100	-	0.6357
Graupel	q_g	500	$\frac{\pi\rho_g}{6}$	3	442	0.89	0	$max\left[10^4, \min\left(\frac{200}{q_g}, 3\times10^6\right)\right]$	0
Cloud Ice	q_i, N_i	890	$\frac{\pi\rho_i}{6}$	3	1847.5	1	0	$\frac{N\lambda^{\mu+1}}{\Gamma(\mu+1)}$	0
Rain	q_r, N_r	1000	$\frac{\pi\rho_w}{6}$	3	4854	1	195	$\frac{N\lambda^{\mu+1}}{\Gamma(\mu+1)}$	0
Cloud Water	q_c	1000	$\frac{\pi\rho_w}{6}$	3	0.316946 $\times10^8$	2	-	$\frac{N_c\lambda^{\mu+1}}{\Gamma(\mu+1)}$	$\min(15, \frac{10^9}{N_c}+2)$

44

Table 2. Morrison mass-size relationship ($m = \alpha D^{\beta}$), terminal velocity-size relationship ($v = [\rho_0/\rho]^{0.54} cD^d$), and gamma PSD parameter values used for each species. N_c used for cloud water μ and N_0 calculations is a constant 100 cm^{-3}. Cloud water μ is calculated as a function of N_c using an empirical relationship described in Martin et al. (1994).

		Morrison *m-D, v-D,* and PSDMSD parameters						
Species	Prognostic Variables	ρ [kg m^{-3}]	α	β	*c*	*d*	N_0 [m^{-4}]	μ
Snow	q_s, N_s	100	$\dfrac{\pi\rho_s}{6}$	3	11.72	0.41	$\dfrac{N\lambda^{\mu+1}}{\Gamma(\mu+1)}$	0
Graupel	q_g, N_g	400	$\dfrac{\pi\rho_g}{6}$	3	19.3	0.37	$\dfrac{N\lambda^{\mu+1}}{\Gamma(\mu+1)}$	0
Cloud Ice	q_i, N_i	500	$\dfrac{\pi\rho_i}{6}$	3	700	1	$\dfrac{N\lambda^{\mu+1}}{\Gamma(\mu+1)}$	0
Rain	q_r, N_r	997	$\dfrac{\pi\rho_w}{6}$	3	841.99667	0.8	$\dfrac{N\lambda^{\mu+1}}{\Gamma(\mu+1)}$	0
Cloud Water	q_c	997	$\dfrac{\pi\rho_w}{6}$	3	3×10^7	2	$\dfrac{N_c\lambda^{\mu+1}}{\Gamma(\mu+1)}$	-

5

Table 3. FSBM mass-size relationship ($m = \alpha D^{\beta}$) parameters and terminal velocity ranges as a function of size (v-D). Note that the snow density varies by bin and thus a range is given. Ranges are also given for all v-D relationships. The subscript k for prognostic variables refers to the kth size bin for a given species.

		FSBM m-D parameters v-D relationships			
Species	Prognostic Variables	ρ [kg m^{-3}]	α	β	v-D
Snow	$q_{s,k}$, $N_{s,k}$	Decreases from 900 kg m^{-3} at D=4.14 µm to 35 kg m^{-3} at D=19.87 mm	$\dfrac{\pi \rho_{s,k}}{6}$	3	Increases from 2×10^{-4} m s^{-1} at D=4.14 µm to 1.4 m s^{-1} at D=19.87 mm~~11.72~~
Graupel	$q_{g,k}$, $N_{g,k}$	400	$\dfrac{\pi \rho_{g}}{6}$	3	Increases from 3.9×10^{-4} m s^{-1} at D=5.43 µm to 4 m s^{-1} at D=8.82 mm~~19.3~~
Liquid	$q_{r,k}$, $N_{r,k}$	1000	$\dfrac{\pi \rho_{w}}{6}$	3	Increases from 5×10^{-4} m s^{-1} at D=4 µm to 9 m s^{-1} at D=6.5 mm

5
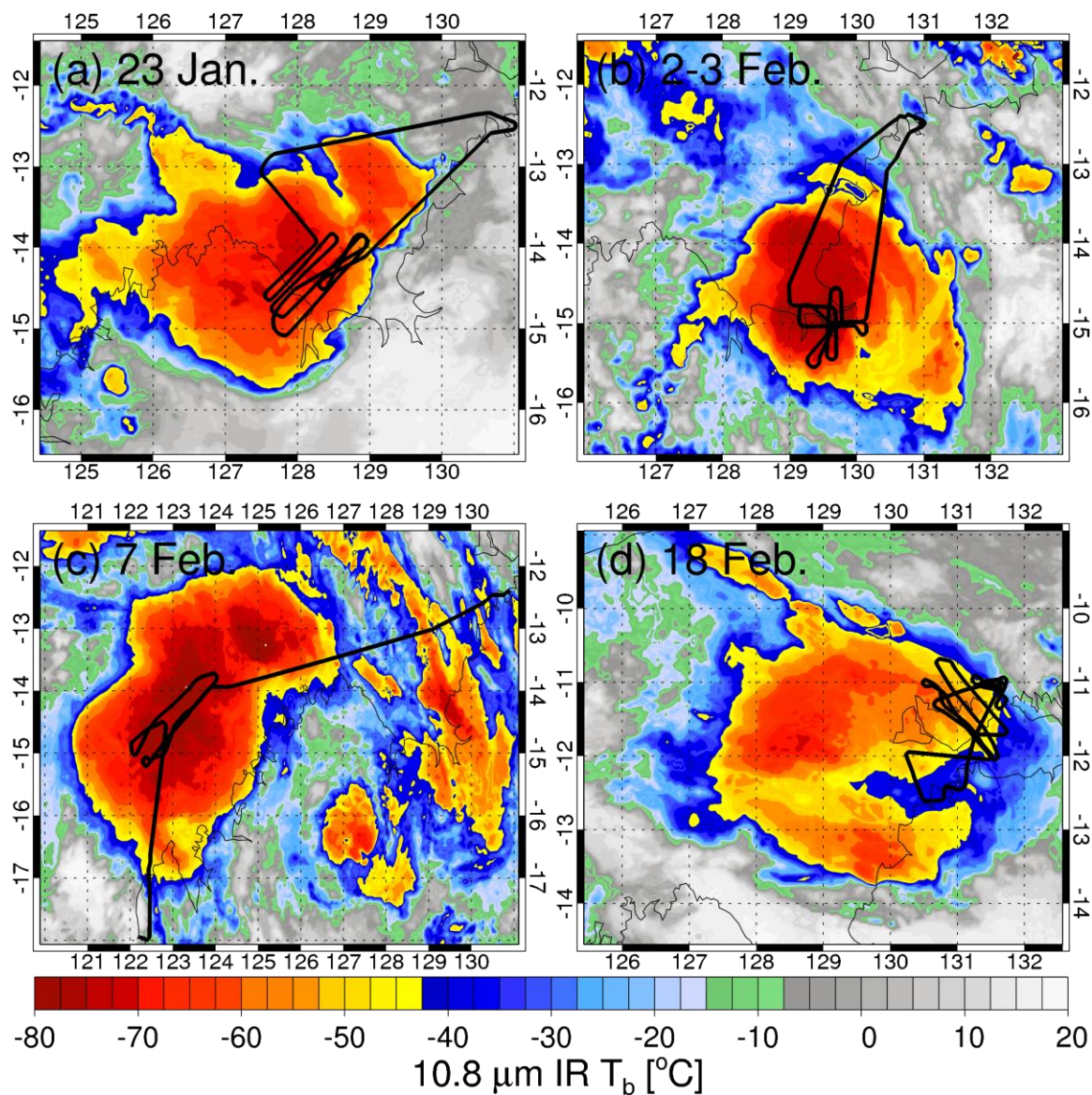
Figure 1: Flight tracks (black lines) for (a) Flight 6 on 23 January 2014, (b) Flight 13 on 3 February 2014, (c) Flight 16 on 7 February 2014, and (d) Flight 23 on 18 February 2014 overlaid on IR imagery from MTSAT-1R during flight sampling. Note that both Flights 12 and 13 flew through the same system (2-3 February), but only Flight 13 is shown.

47
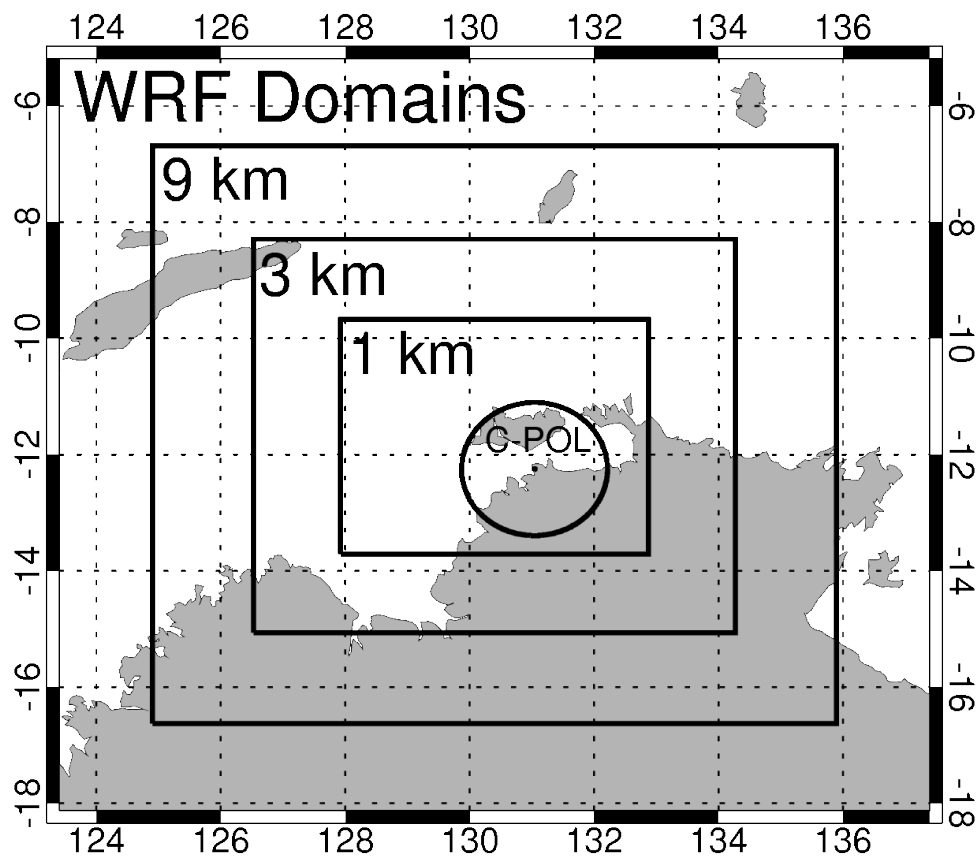
Figure 2. The WRF domains used for the 18 February 2014 simulation. The circle indicates the 150-km range ring of the C-POL radar located at Gunn Point.
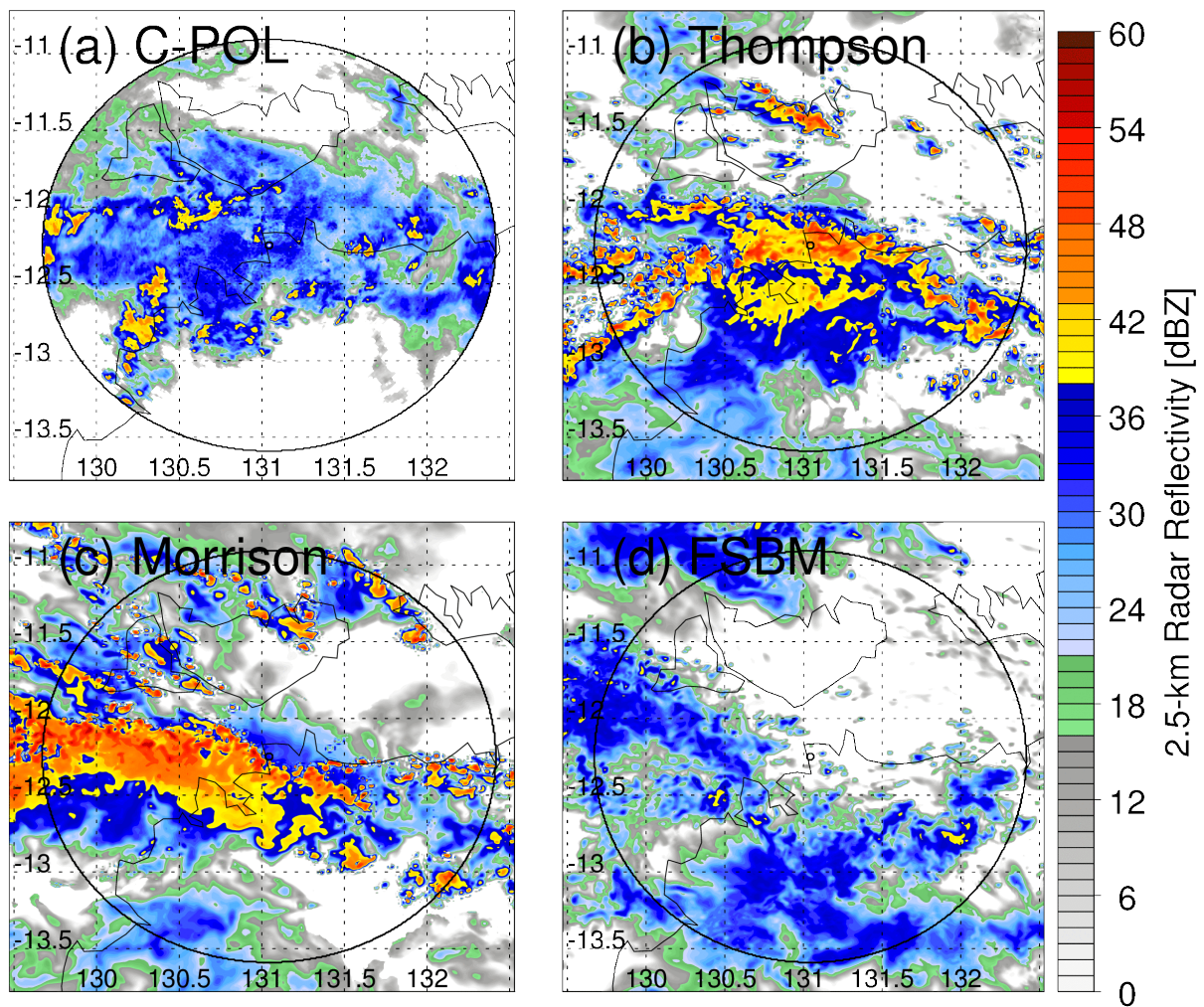
Figure 3. Representative 2.5-km altitude horizontal radar reflectivity cross-sections on 18 February 2014 at 18Z for (a) C-POL observations, (b) Thompson, (c) Morrison, and (d) FSBM schemes. The circle indicates the 150-km range ring of the C-POL radar.
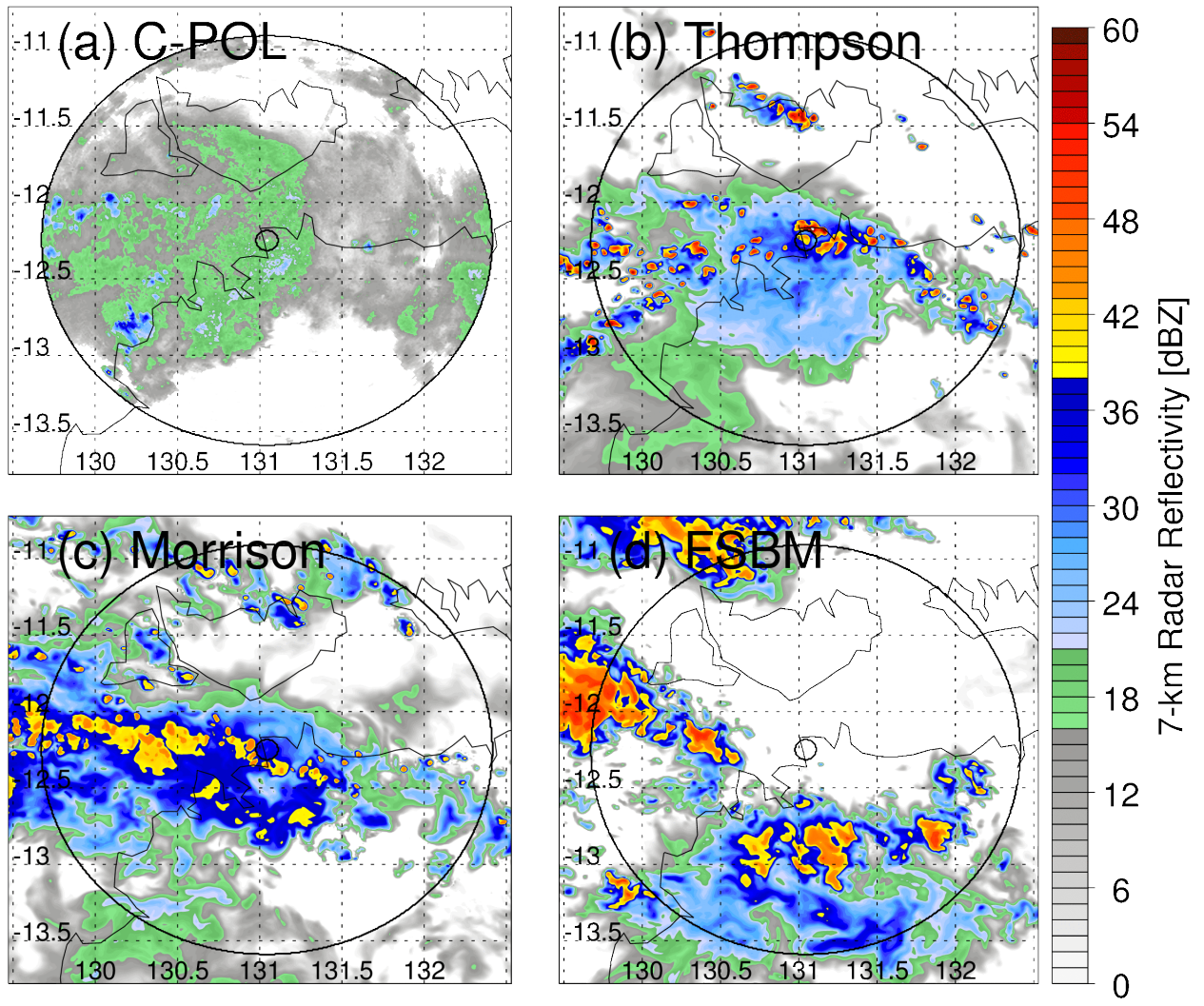
Figure 4. As in Fig. 3, but for 7-km altitude.
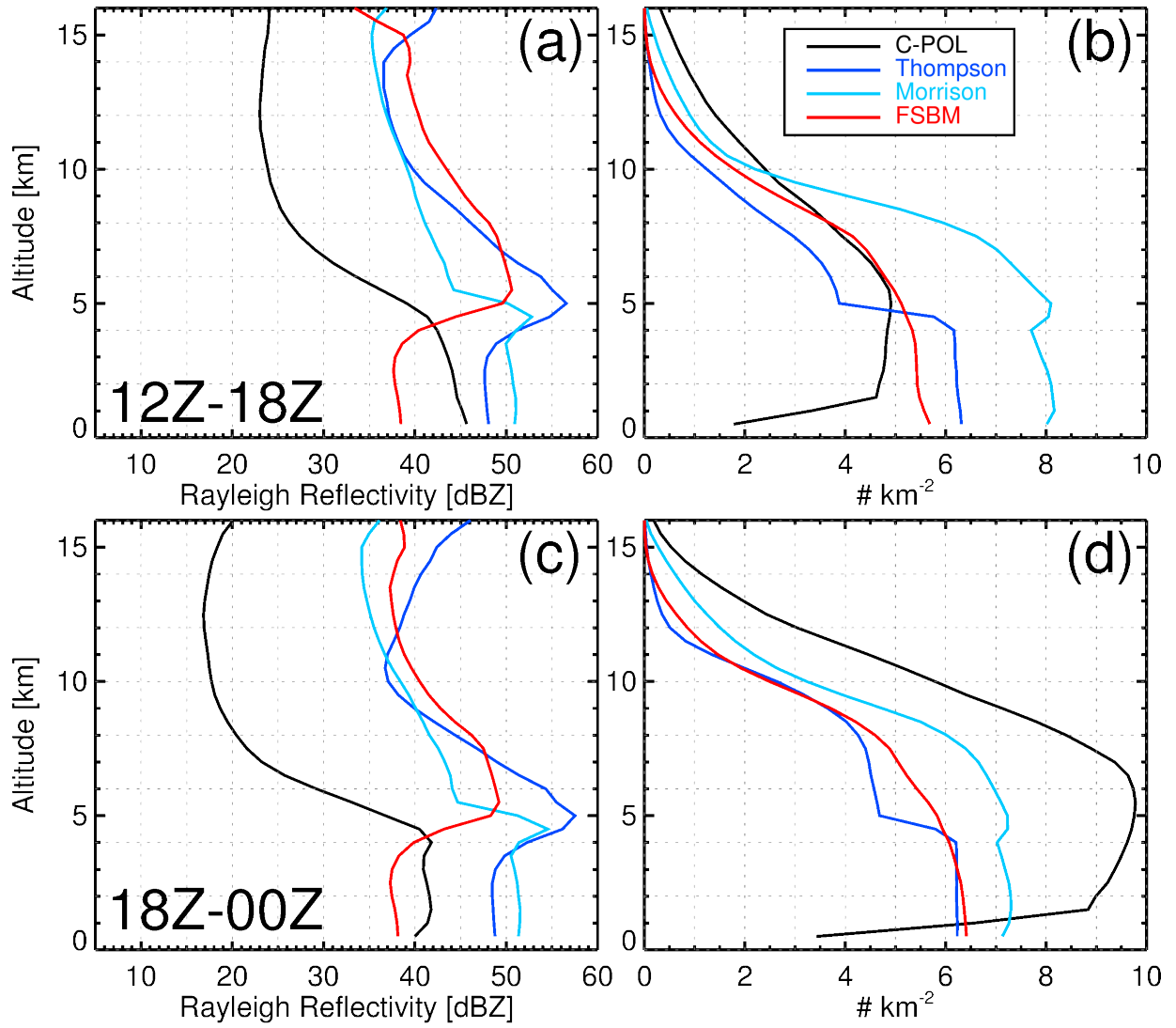
Figure 5. Rayleigh reflectivity 99th percentile profiles for the observed and simulated 18 February 2014 MCS case during (a) 12Z-18Z on the 18th and (c) 18Z on the 18th to 00Z on the 19th. Sample sizes normalized by domain area are shown in (b) and (d) for 12Z-18Z and 18Z-00Z, respectively at 16Z. Only reflectivity values > 5 dBZ are included. The 90th percentiles are shown in (a) and the 99th percentiles are shown in (b). Only reflectivity values > 5 dBZ are included. Retrieved C-POL reflectivity is in black, Thompson in dark blue, Morrison in cyan, and FSBM in red.
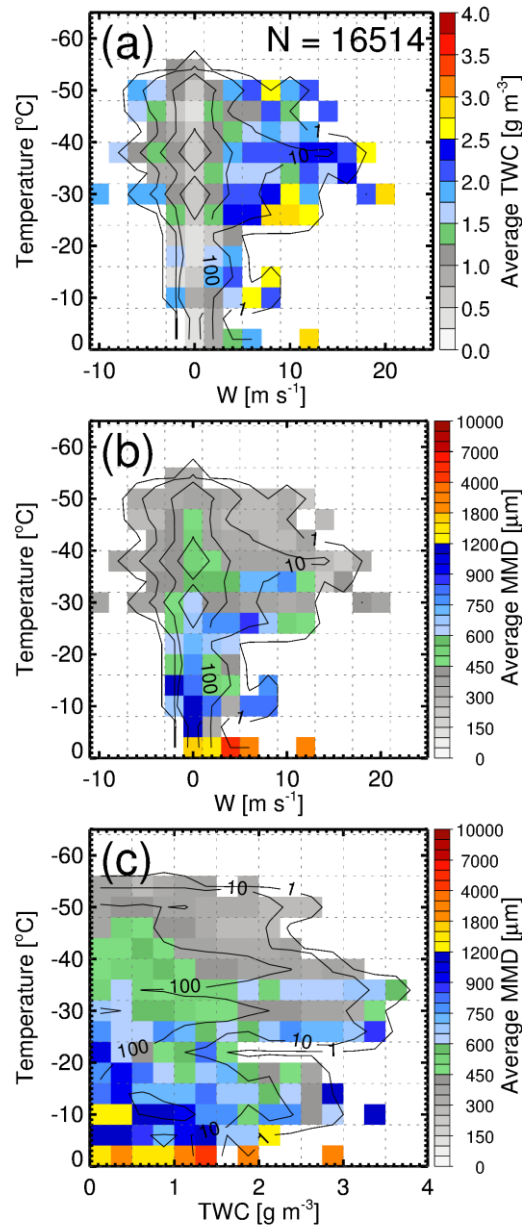
Figure 6. Observed joint histograms using Darwin HAIC-HIWC data: (a) w-T bins color-filled with average TWC, (b) w-T bins color-filled with average MMD, and (c) TWC-T bins color-filled with average MMD. Observational sample size is shown in the top right corner and order of magnitude sample sizes are contoured in black. Bin sizes are 2~~1.5~~ m s^{-1} for w, 0.25 g m^{-3} for TWC, and 4 °C for temperature. Note that the average MMD color-scale is nonlinear and accentuates variability for average MMDs < 1 mm.
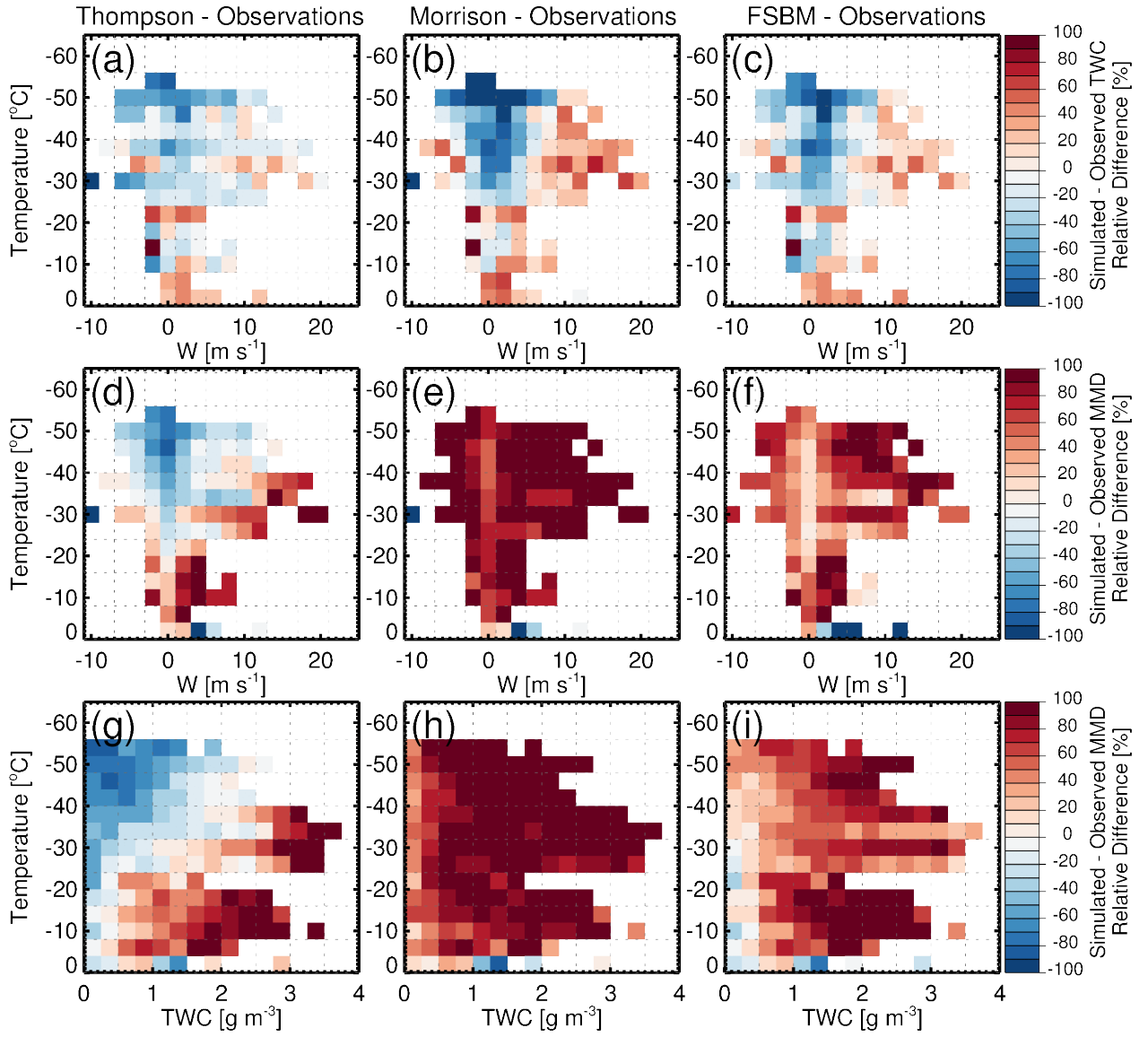
Figure 7. Joint histograms color-filled with relative differences between observed and simulated (a)-(c) TWC as a function of *w*-T bins, (d)-(f) MMD as a function of *w*-T bins, and (g)-(i) MMD as a function of TWC-T bins. Differences for the Thompson scheme are shown in the left column, the Morrison scheme in middle column, and the FSBM scheme in right column. Observed values are subtracted from simulated values, and only bins where observational data are available are shown.

53
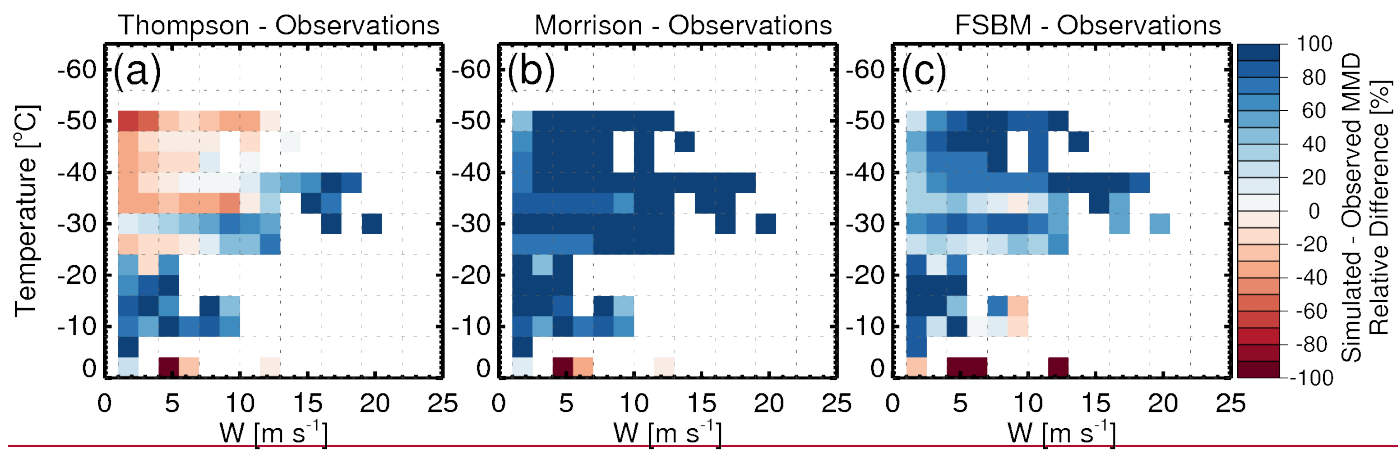
Figure 8. As in Fig. 7, but for observed MMDs subtracted from simulated MMDs.
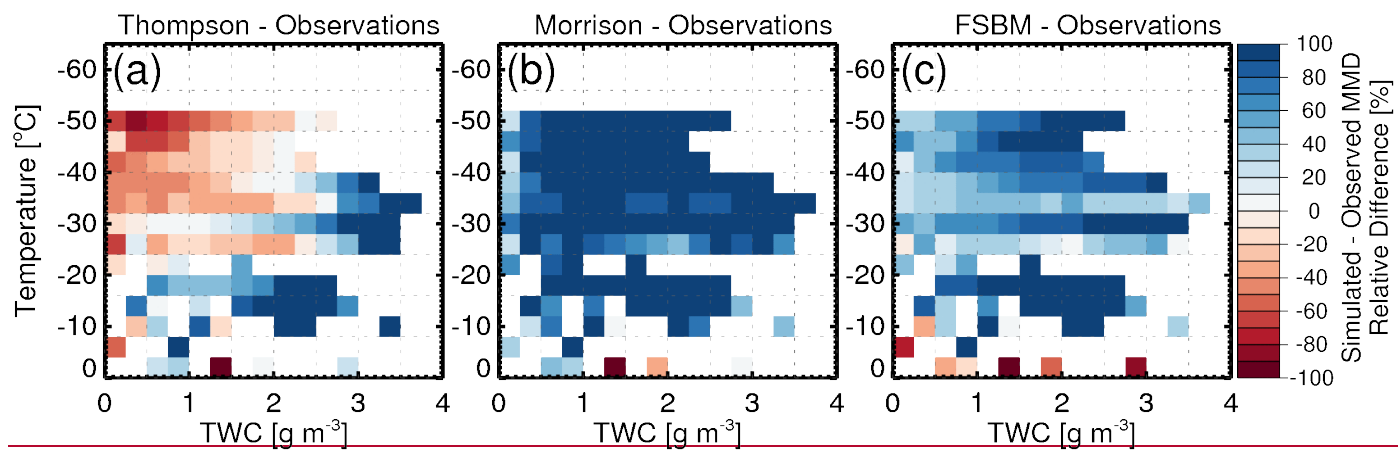
Figure 9. As in Fig. 8, but the abscissa is TWC bins.
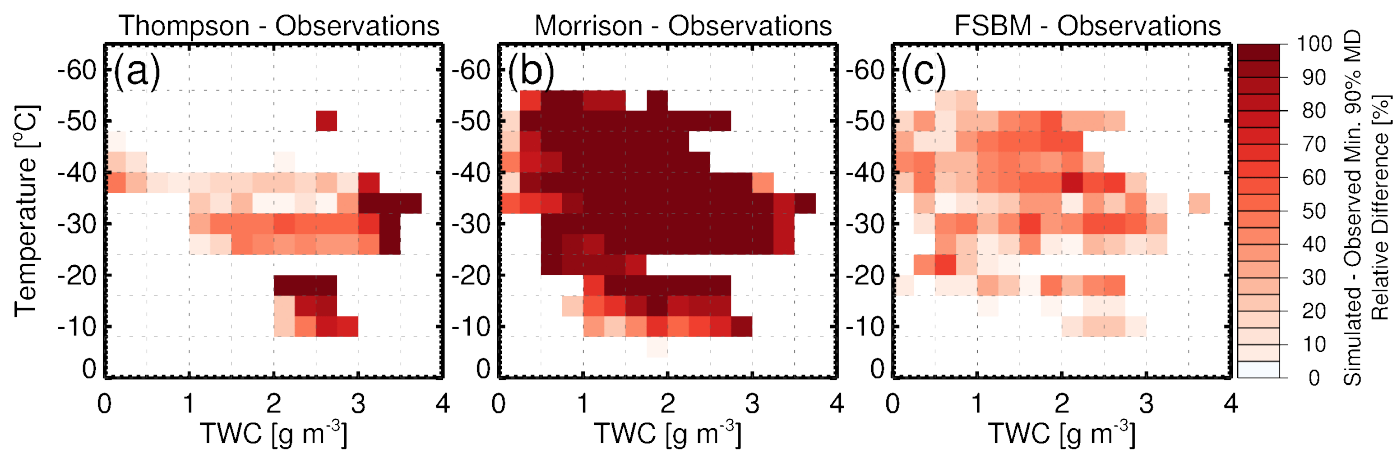
Figure 8. Joint histograms of positive relative differences between observed and simulated minimum 90% MDs as a function of TWC-T bins for (a) Thompson, (b) Morrison, and (c) FSBM schemes.
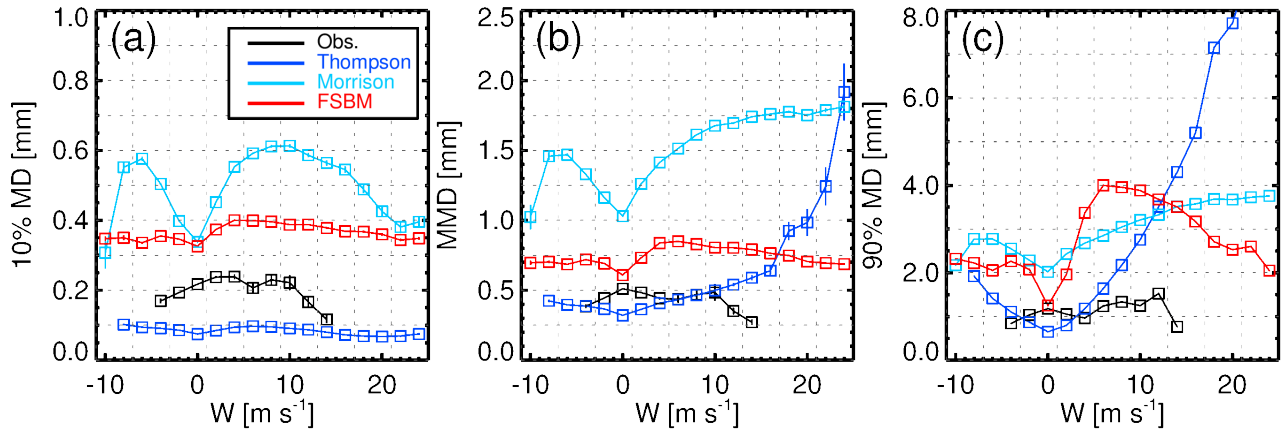
Figure 9. Average percentile mass diameters as a function of w bins (2 m s[-1] bin widths) for temperatures between -32 °C and -40 °C. 10% MDs are shown in (a), MMDs in (b), and 90% MDs in (c) with observations in black, Thompson in dark blue, Morrison in cyan, and FSBM in red. Vertical bars extending from each box represent standard errors; however, note that standard errors are rather small due to relatively large sample sizes.
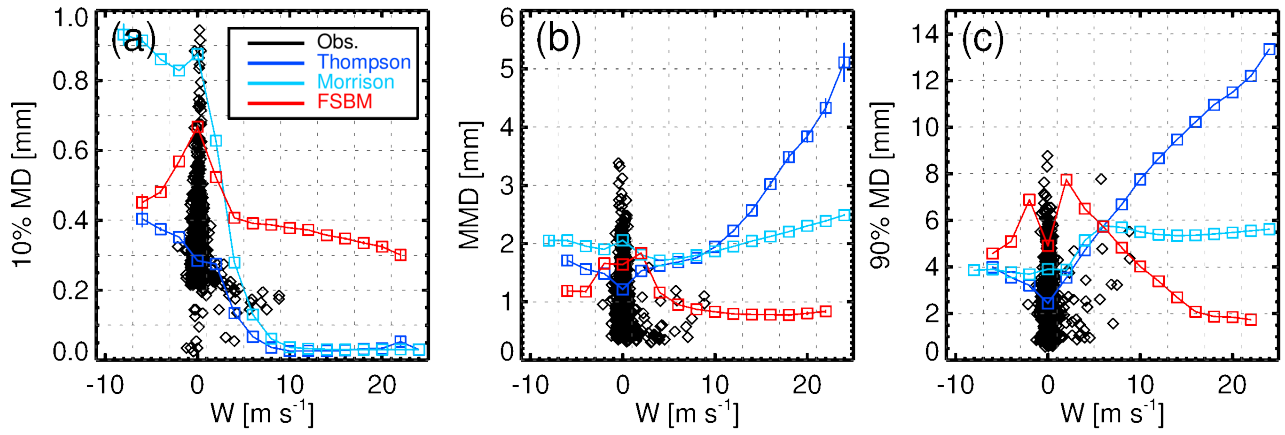
Figure 10. As in Fig. 11, but for temperatures between -8 °C and -16 °C. Observations are plotted as black diamonds displaying individual data points due to lower sample sizes in this temperature range.
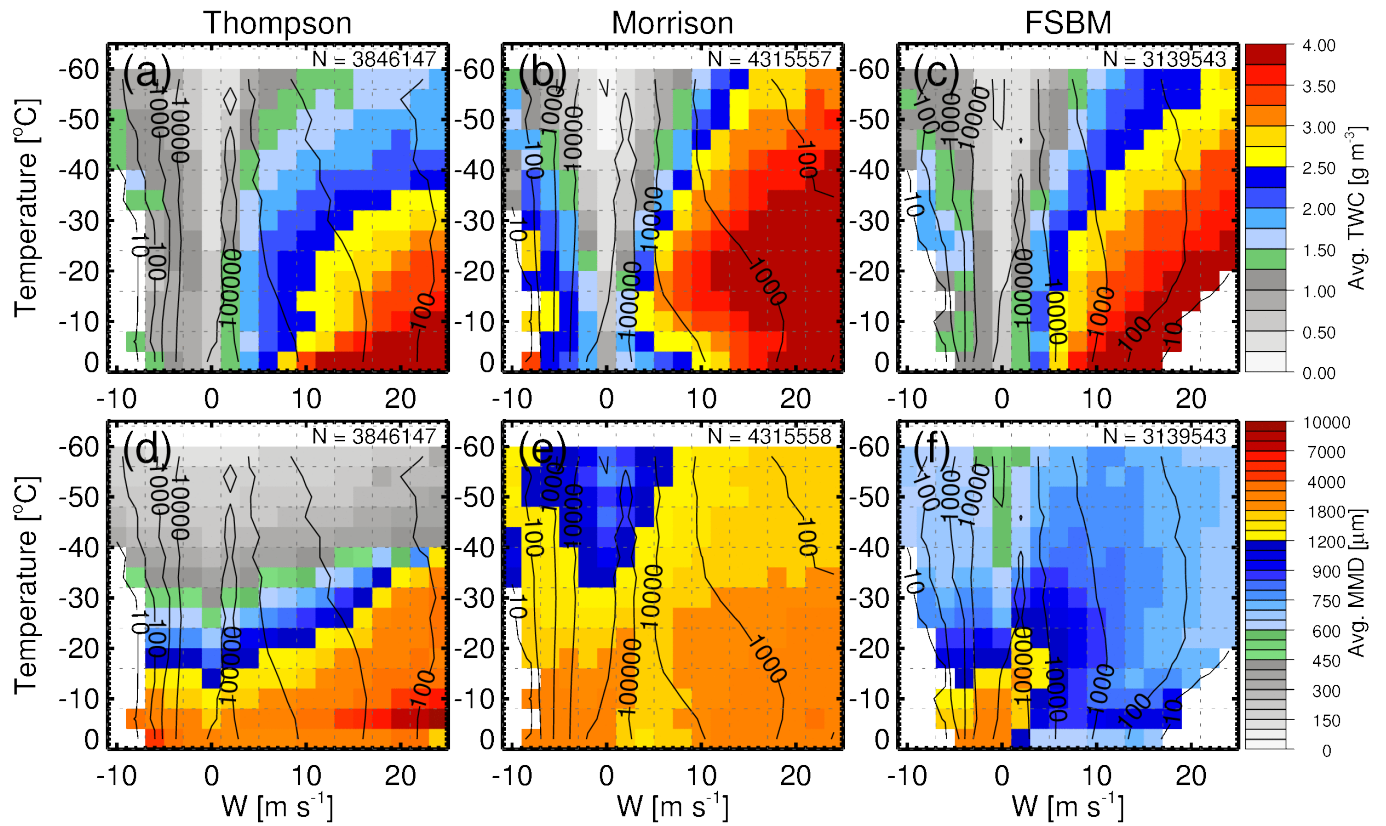
Figure 11. Joint histograms of w and temperature below freezing using bins sizes of ~~1.5~~ 2 m s^{-1} for w and 4 °C for temperature. Color-fill is average TWC for (a)-(c) and average combined-hydrometeor MMD for (d)-(f). The Thompson scheme is shown in (a) and (d), the Morrison scheme in (b) and (e), and the FSBM scheme in (c) and (f). Sample sizes are shown in the upper right corner, and order of magnitude sample sizes are contoured in black.
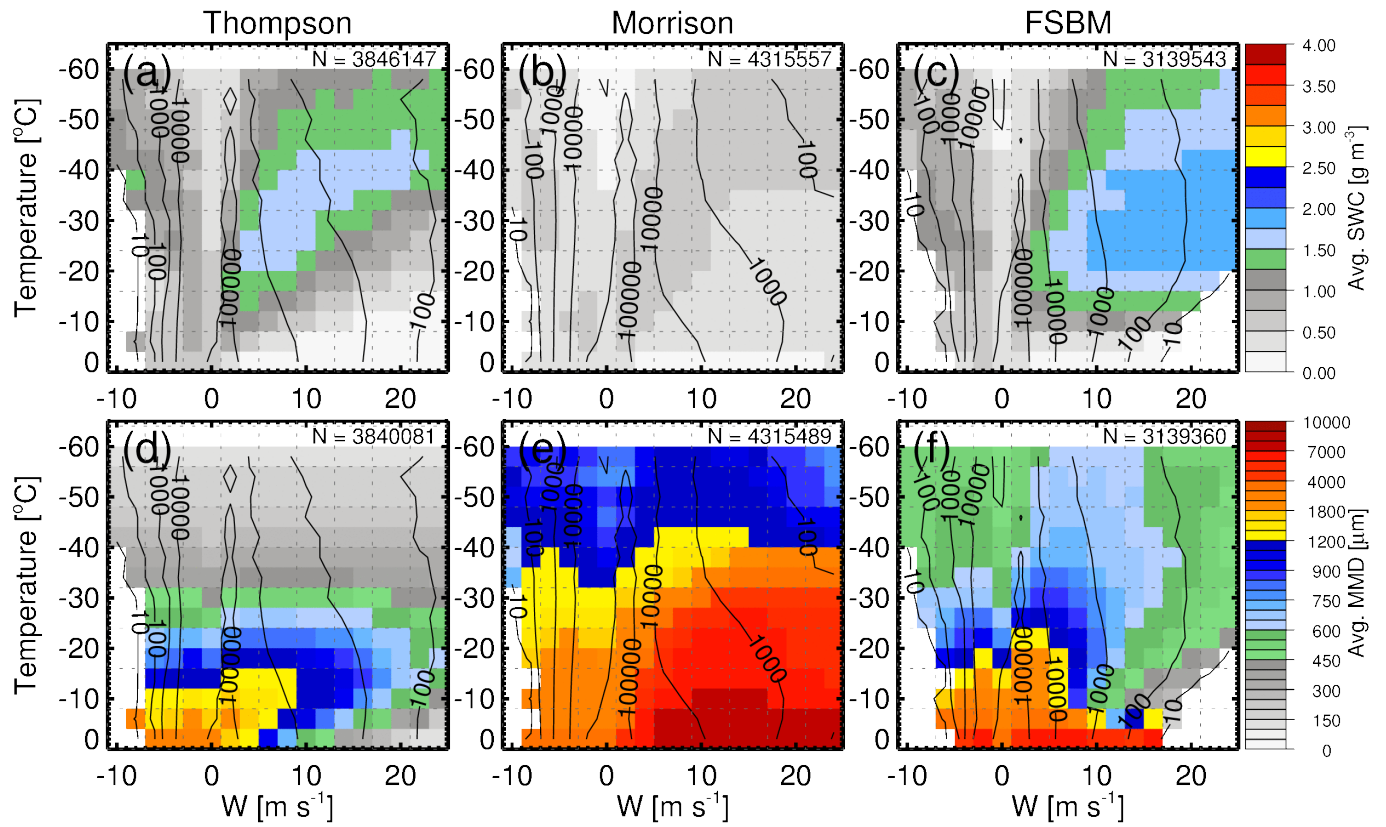
Figure 12. As in Fig. 1~~1~~3, but for snow water content (SWC) color-filled in (a)-(c) and average snow MMDs color-filled in (d)-(f).
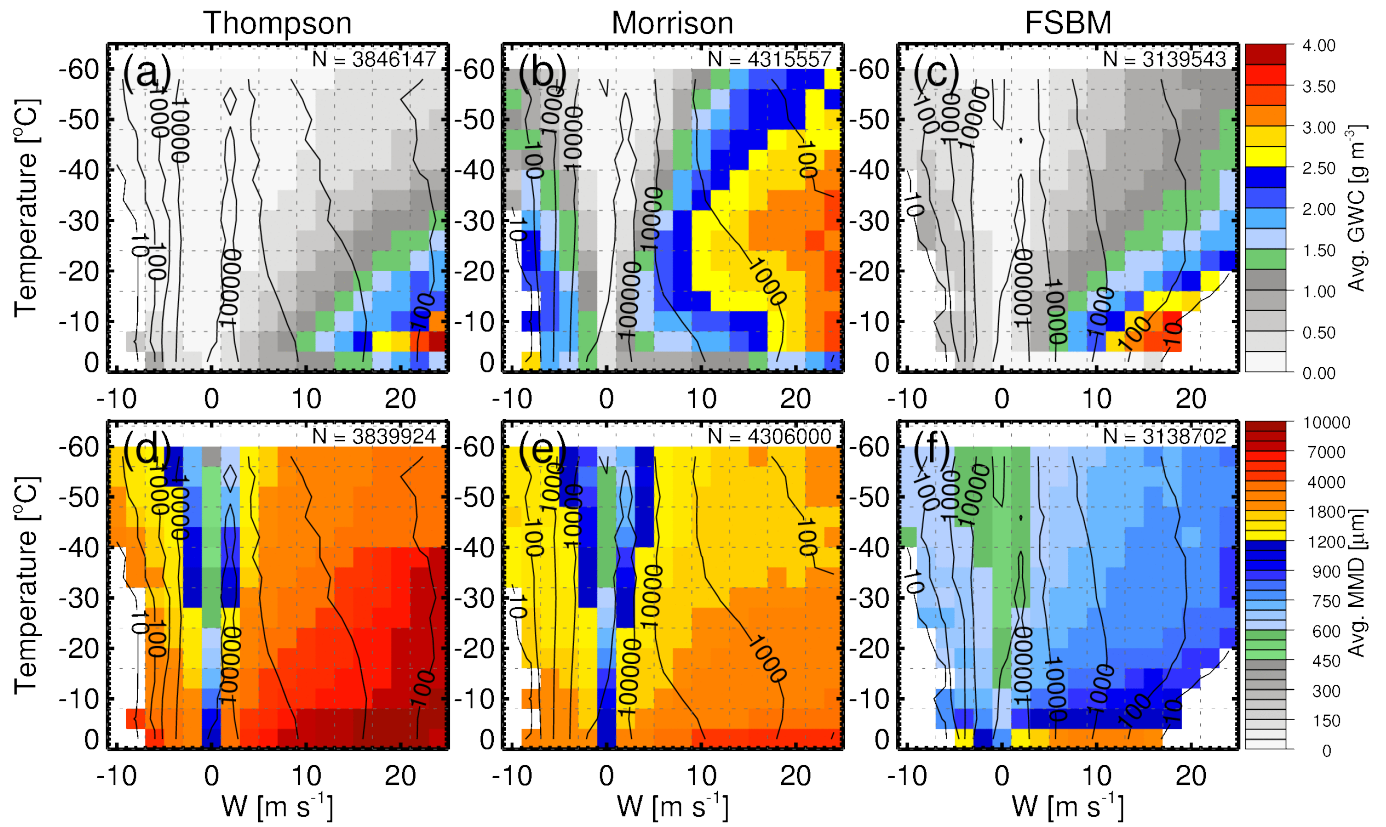
Figure 13. As in Fig. 1~~1~~3, but for graupel water content (GWC) color-filled in (a)-(c) and average graupel MMDs color-filled in (d)-(f).
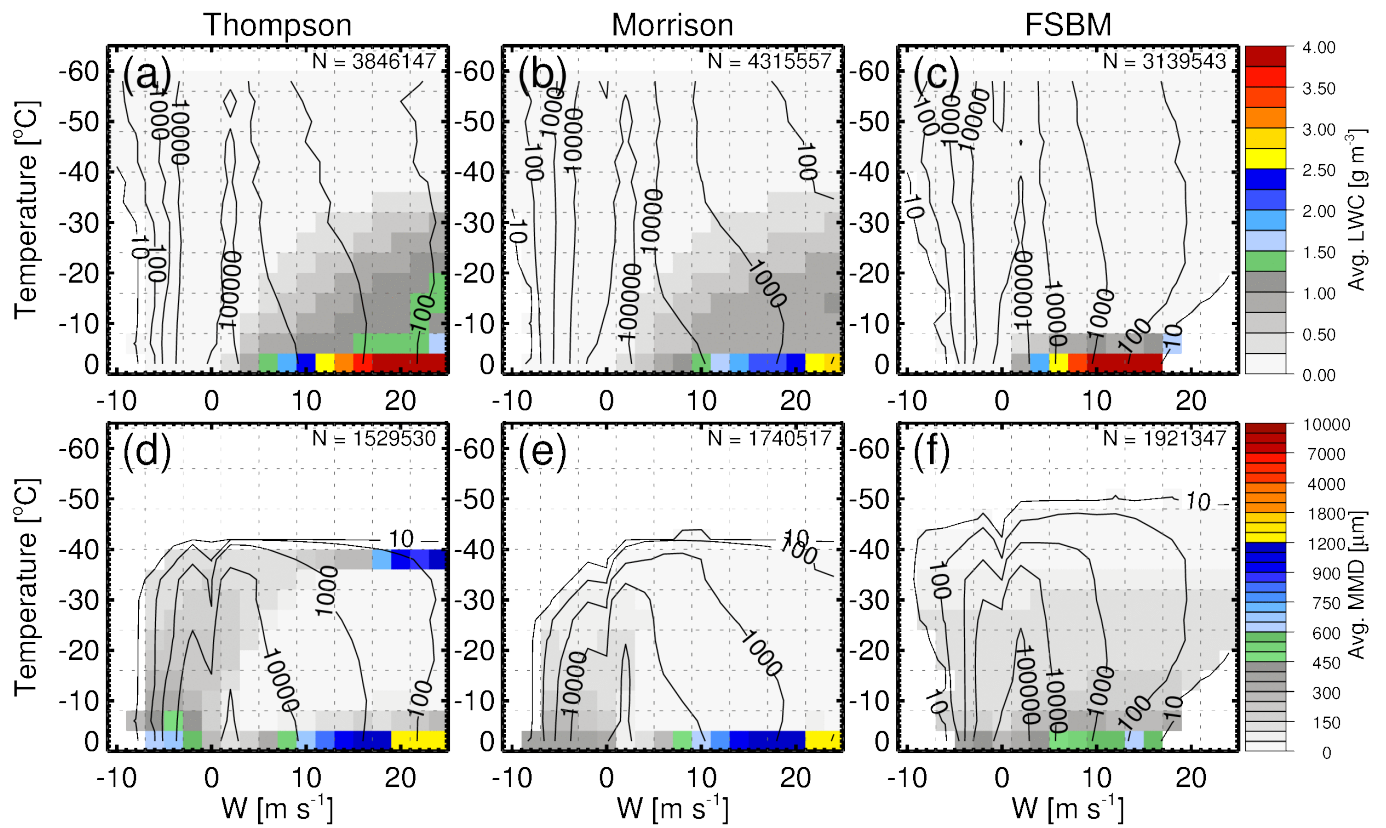
Figure 14. As in Fig. 131, but for liquid water content (LWC) color-filled in (a)-(c) and average liquid MMDs color-filled in (d)-(f).
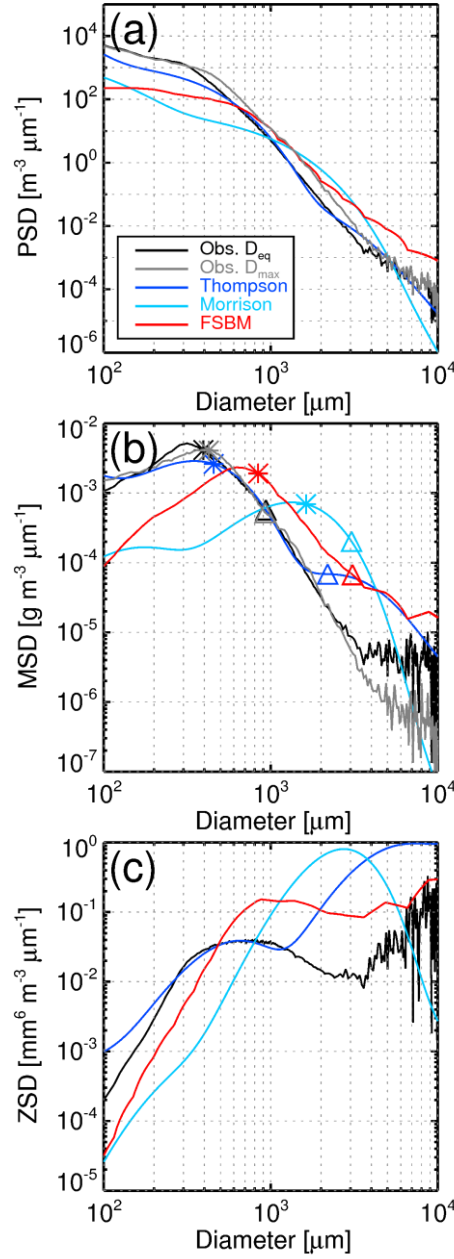
Figure 15. Observed and simulated composite average (a) PSDs, (b) MSDs, and (c) ZSDs for TWC between 2 and 2.5 g m^{-3} and temperatures between -32 °C and -40 °C, where TWC is the integrated mass in diameters less than the observational maximum of 12.85 mm. Observed distributions for the D_{eq} diameter definition are in black and for the D_{max} definition are in grey. Thompson is in dark blue, Morrison in cyan, and FSBM in red. Asterisks and triangles overplotted on the MSDs in (b) are the MMD and 90% MD, respectively. Note that both axes are logarithmic.