Responses to the comments of Referee #1

Referee #1: The paper reports on a regional simulation of an MCS where large ice water contents were observed. Four different microphysical schemes were compared to observations. The paper is clear and to the point. I think that only minor revisions are necessary before publishing.

Response: We would like to express our acknowledgement for your efforts and constructive comments. Our point-by-point responses are given below. For convenience, the reviewers' comments are in **black** fonts, and our point-by-point responses are in **blue**.

This is obviously a nice test case and one that others may try to reproduce themselves. The observations are freely available, but will scripts to reproduce the sampling outlined here also be available?

Response: The scripts used in our study do not have detailed comments currently, so someone who is interested in this study and wants to reproduce the sampling can contact us directly, and we will give some instructions how to use the scripts.

Is there an accepted definition for HIWC region. Can this be indicated and compared in some way to show that models have some skill at predicting these regions? Including something like this would link the results nicely to the operational motivation for this work.

Response: The broad definition of a HIWC region is "Regions with high ice water content (HIWC), composed of mainly small ice crystals, frequently occur over convective clouds in the tropics. Such regions can have median mass diameters (MMDs) < 300 μm and equivalent radar reflectivities < 20 dBZ". However, currently there is no unified definition for the threshold of IWC, and the threshold of 1 g cm^{-3} is used in our study. From the simulations in this study, regions with higher IWC can be predicted by the model, while reflectivities of these regions are overestimated. It indicates that model can capture the regions with higher IWC, but it cannot represent the phenomenon of numerous small ice crystals contributing to the IWC. This is summarized in the last section.

line 55. Could refer to Keinert et al. who have carried out laboratory experiments to investigate droplet freezing/shattering in this temperature range. (https://www.researchgate.net/publication/343566907_Secondary_Ice_Production_upon_Freezing_of_Freely_Falling_Drizzle_Droplets)

Response: We have cited the paper in our revised manuscript.

line 159. Feel free to ignore this because there are always more that can be added to an intercomparison, but given the operational importance of HIWC events I'm surprised that the Thompson scheme was also not included in the mix - I believe it is operationally used (or was) in the NOAA RUC model.

Response: We selected microphysics schemes for study that parameterized ice species differently. WSM6 is a single-moment scheme for cloud ice, snow and graupel; the Morrison scheme is a double-moment scheme for cloud ice, snow, and graupel; P3 represents ice species very differently. The Thompson scheme uses a double moment for cloud ice and a single moment for snow and graupel. If possible, we will add the Thompson scheme in our future modeling studies.

line 182 - '...likely associated with..' - i think you should be able to say yes or no to this by inspecting era data rather than leaving it hanging.

Response: We have revised the statement and removed the word "likely".

line 189 'using the assumptions consistent' - does this include the shape of the psd or just the mass-size/density assumptions?

Response: Yes, it includes the parameters of the PSDs in addition to mass-size/density assumptions. We have added a description "including the characteristics of the cloud species and PSDs" in our revised manuscript to clarify this point.

line 207-210. Should be careful to add that your statement is for this metric: BT. e.g. this means deep convective areas as defined by this BT metric are larger in MORR than.... You could define it by updraft or specific humidity...

Response: We have revised the statement to make it clear that it is for the BT metric, which now reads "This means deep convective areas as defined by BT metric are larger in MORR than the observations at an early stage in the system."

line 230. it may be too messy but it might be worth trying to add the cumulative frequency contours from the obs to the model panels to provide an easier way to compare across?

Response: The white contours in Fig. 7 represent cumulative frequency. We have revised the description to make this more clear.

line 250. even though there is a bias in sampling - is this not compensated with the BT sampling methodology?

Response: Yes, the BT sampling method can compensate for part of the sampling bias. We have revised the related sentences in our manuscript, which now reads "It should be noted that the NRC Convair 580 operations avoided the cloud regions with high reflectivity due to safety regulations, and thus it did not approach high reflectivity regions (red zones on the pilot's radar) within 30 nautical miles (~55.56 km). However, the BT sampling method has been used to minimize these aircraft sampling biases."

line 252. can you estimate the impact? If not i think you have to assume its unbiased...?

Response: We cannot estimate the impact. But, we have revised the manuscript to emphasize that the BT sampling method minimizes the sampling biases, which now reads "However, the BT sampling method has been used to minimize these aircraft sampling biases."

line 284. how is psd spread quantified?

Response: The "spread" here means the range between the 25th and 75th percentiles in Fig. 10b. We have indicated this in our revised manuscript, which now reads "…same PSD spread (range between the 25th and 75th percentiles)…".

line 302. reasonably to within x% ?

Response: We have updated the related statement, which now reads "Generally all the simulations, especially MORR, reproduce the TWC reasonably within the same order of magnitude as observations at the three temperature levels, with biases within 38% at −10 ◦C"

line 311. Possibly add a radar weighted mean size to link to radar results?

Response: We should clarify that observed radar reflectivity profiles were sampled along the flight track of NRC Convair 580, whose horizontal locations are different from those of observed TWC/PSD samples from SAFIRE Falcon 20 (Fig. 1). Part of observed TWC/PSD samples at −10 ºC and all observed TWC/PSD samples at −45 and −30 ºC are from SAFIRE Falcon 20 (Fig. 1), implying that observed TWC/PSD at −45 and −30 ºC may be not fully consistent with observed radar reflectivity in the upper levels in this study. Based on this bias, we cannot link the radar weighted mean size to radar results directly.

line 333. 'it is found that the main microphysical process rates at -45 and -30C are the same as those within profiles containing HIWC regions at -10C'. I don't think you mean this but i'm interpreting it as using process rates at -10C as a proxy for what is going on at -45 and -30C. If so, then i would expect processes involving graupel production to be different between -10C and -45C. Additionally, in strong convection, the loss of liquid at lower levels controls the liquid being transported to higher up and the eventual anvil evolution. At -45C the freezing is dominated by homogeneous freezing, whereas at -10C it will be heterogeneous freezing or secondary ice production. Therefore i don't think you can use the process rates at -10C as proxy for -45C.

Response: Our original description confused both reviewers so we have reworded this. We meant that the main microphysical processes are similar at the same vertical levels when comparing sampling profiles with TWC > 1 g m^{-3} at −10 ºC to sampling profiles with TWC > 0.1 g m^{-3} at −45 and −30 ºC. We have revised the manuscript to make this more clear, which now reads "Through comparing between simulated sampling profiles with TWC > 0.1 g m^{-3} at −45 and −30 ◦C and simulated sampling profiles with TWC > 1 g m^{-3} at −10 ◦C, it is found that their main microphysical processes are the same at the same vertical levels."

line 335. i think i disagree here. Transporting more cloud droplets to homogeneous freezing altitudes/temperatures could lead to more numerous small ice crystals.

Response: We have deleted this sentence as we agree it is not fully validated by the available data.

line 349-350 '...substantially underpredict the ice particle number for 0.1 mm < Dmax < 3 mm and overpredict the vertical motion in the HIWC regions, which results in stronger and higher-extended simulated radar reflectivity...' and line 355-356 '...an underestimate of ice particle number concentration, especially graupel, leads to large reflectivities...'

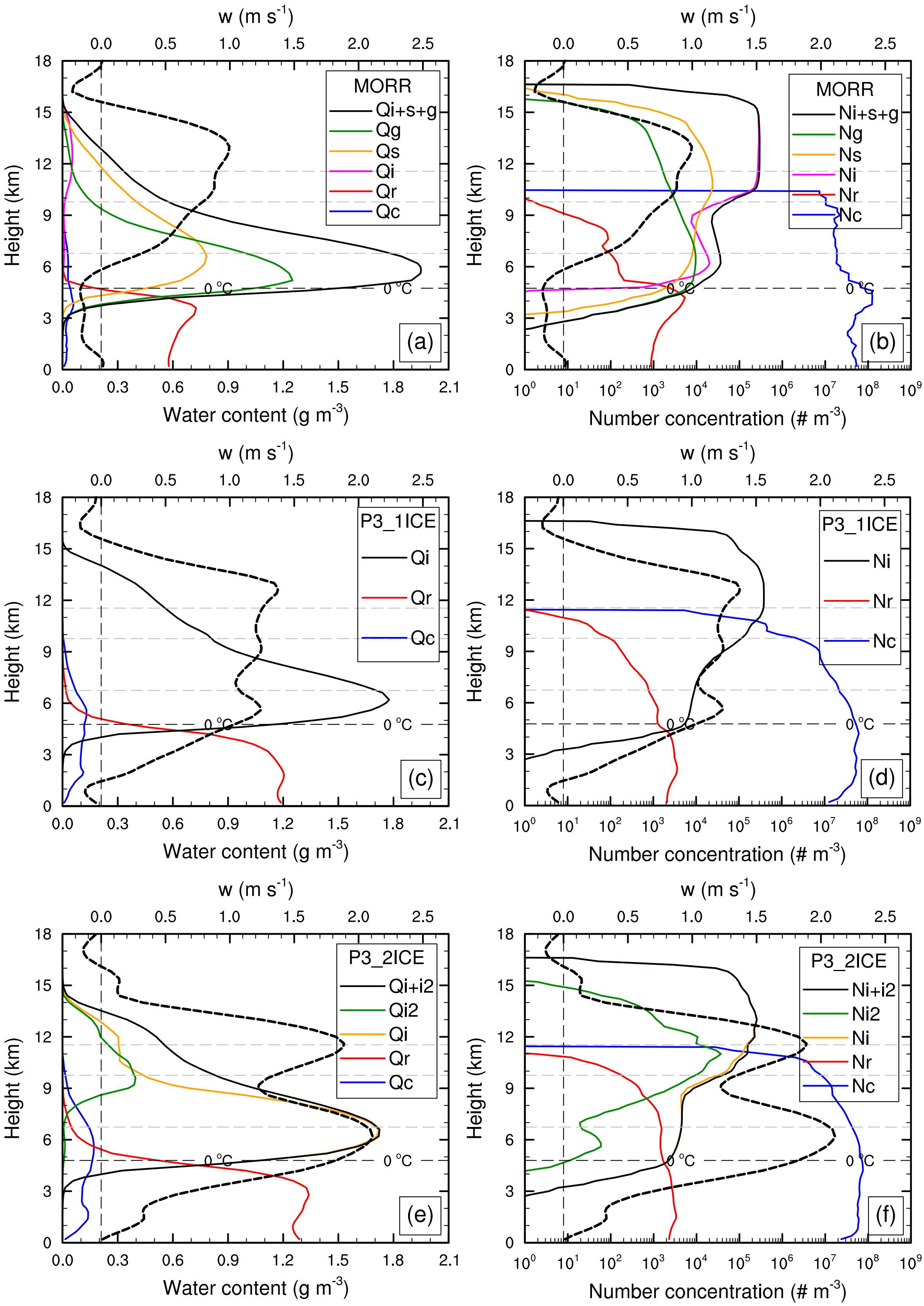
Because the psd is a gamma or exponential distribution in the model i accept that reducing total number concentration will lead to an increase in the reflectivity for a fixed mass. But invoking that strong correlation between number concentration dominated by the small end of the PSD and the large end of the PSD that affects the radar reflectivity is not a given for the real world. In the real world, underpredicting number concentration will not necessarily be the direct source of the radar reflectivity overestimate - its the SUM_i(m_iˆ2) integral where i is the i_th size bin. If the total mass is correct the overestimate of Ze must come from the mass being in too large size bins as your PSD comparisons suggest.

I can see that in the appendix the radar reflectivity is formulated to depend upon Nt due to the gamma assumption, but it feels more physical to relate the effect to the large end of the PSD.

Response: From the model perspective, if the total mass is correct, the underestimate of total number concentration will lead to a larger mean-mass diameter with a gamma or exponential distribution assumption. This implies that mass will be concentrated at the large end of the PSD resulting in an overestimate of reflectivity. We agree that in the real world it is more physical to relate the reflectivity to the large end of the PSD. We have revised the related statements based on this point which now reads "an underestimate of ice particle number concentration, especially graupel, leads to mass concentrating at the large end of PSD resulting in large reflectivities" and "Therefore, the collection of cloud water by graupel is the key source term of total IWC at −10 ◦C in MORR, which increases the mean mass/size of graupel and does not directly influence its number leading to mass concentrating at the large end of PSD."

line 354 - do you need to add a total ice category line to the MORR plot to compare to P3?

Response: Figure 12 has been updated.



line 371. you could resample the data to match the liquid water content from the model and observations and then see if the ice properties etc are biased.

Response: To minimize the location biases of samples, the BT sampling method was adopted, which is the best method we are aware of. There are few samples simultaneously satisfying the BT and liquid water content conditions. We have revised the manuscript accordingly, which now reads "It should be noted that both the NRC Convair 580 and SAFIRE Falcon 20 avoided cloud regions with strong updrafts where presence of liquid phase is expected, however, the BT sampling method has been used to minimize these sampling biases."

line 375. ice nucleation = homogeneous freezing? or hom+het freezing?

Response: The "ice nucleation" in Morrison scheme considers nucleation of cloud ice from homogeneous and heterogeneous freezing on aerosol in deposition/condensation-freezing modes (Morrison and Gettelman 2008). It is a function of temperature following Cooper (1986). We have indicated this in our revised manuscript.

Reference:

Cooper, W. A., 1986: Ice initiation in natural clouds. Precipitation Enhancement–A Scientific Challenge, Meteor. Mongr., 43 , Amer. Meteor. Soc. 29–32. https://doi.org/10.1007/978-1-935704-17-1_4

Morrison, H., & Gettelman, A. (2008). A New Two-Moment Bulk Stratiform Cloud Microphysics Scheme in the Community Atmosphere Model, Version 3 (CAM3). Part I: Description and Numerical Tests, Journal of Climate, 21(15), 3642-3659. https://doi.org/10.1175/2008JCLI2105.1

line 378. i could not really see this figure - all i can see is the total ni_tend going out of range. It looks like a rime splintering secondary ice production is represented but has no effect?

Response: The axis range of Fig. 13d has been updated, so that the processes can be seen more clearly. The rime splintering secondary ice production is activated at air temperatures between −3 and −8 °C, so it is not the dominant microphysical process at −10 °C. However, it would influence the Ni profile due to vertical advection of Ni. This will be discussed in more detail in a forthcoming paper.

line 451. the 3km radar results should appear earlier than the final page i think.

Response: In our opinion, the 3-km radar results are not the main point in this study, and we only used them for our discussion. Therefore, we prefer to keep the original location.

line 459 it doesn't look that clear cut to me. based on red triangles in fig 11 i score it as p3-2ice=2, wsm6=1, morr=1
The psds using the mean of the De metric: -10C wsm6, morr, p3_1, p3_2 -30C p3_1, morr, wsm6, p3_2 -45C p3_1, morr, wsm6, p3_2
cfads - the cumulative curves from p3_1 seem to match best with obs. then wsm6, morr, p3_2
If you are going to say which is best i think you need some quanitative measures to quote.

Response: The reviewer makes a valid point. It is hard to say that any one scheme is better than the other in all the metrics examined in our study. We have revised this statement based on your comment, which now reads "In conclusion, no one microphysics scheme outperforms the other scheme in simulating this tropical oceanic MCS as evident from examining the simulated soundings, brightness temperature, radar reflectivity, ice particle size distributions, total water content and number concentration."

Responses to the comments of Referee #3

Referee #3: The authors propose a study where they compare the ability of 4 different parametrisations of clouds microphysics in WRF, to simulate an extreme weather event during the campaign HAIC-HIWC over Cayenne in May 2015. Attention is focused on the capacity of the 4 schemes to simulate the High Ice Water Content (HIWC) for IWC > 0.1g/m3 with their associated particle size distributions (PSD). Before the presentation of their study, the authors provide a short review of results from former studies on simulation of HIWC and analysis of the HAIC-HIWC dataset. Hence, they highlight that median mass diameter (MMD) for IWC>1g.m3 increase with the temperature and decrease with increasing total water content (TWC); meaning that HIWC at high altitude are made with small ice crystals. Also, it was showed that X-band radar do not have sensitivity to detect HIWC, but it can be countered using polarimetric parameters such specific differential phase (Kdp) and differential radar reflectivity ratio (Zdr) instead of radar reflectivity factors (Z) only. More importantly, the secondary ice production (SIP) process must have greater contribution in HIWC, than ice nucleating particle to explain their high number concentration of ice particles. A new process of "freezing- drop-shattering" has been recently proposed by Korolev et al. 2020. Simulations with single and double moment microphysical schemes of extreme weather event, explored during HAIC-HIWC in Darwin, showed overestimation of Z in C-band. Also, it showed overestimation of MMD for TWC> 1g/m3. Another simulation of an event during HAIC- HIWC in Cayenne, failed to represent observed IWC and PSD, suggesting a lack of the representation of SIP processes. In the presented study, there is no new parametrisation of clouds microphysics tested. But, two parametrisations are tested for the first time to simulate HIWC, where the density of the ice is predicted for one or two ice species (called in the study P3-1ICE and P3-2ICE). The other microphysical scheme are a single moment with 6 species of hydrometeors (WSM6) and a double moment with 5 hydrometeors types (MORR). Quality of simulated HIWC and PSD is tested for each scheme, through the comparison with observed Brightness temperature (Tb) of the channel 4 of GOES-13, the observed radar reflectivity factors (Z) in the X-band and the observed PSD. The authors conclude a good prediction in average of temperature, dew point and winds fields, where prediction with MORR are closer to observations. The comparison of observed brightness temperature with the simulations shows better results with the scheme P3-2ICE. However, simulated Z with the four microphysical scheme are larger than the Z observed. The four microphysical schemes give bad representation of particle size distributions when compared with the observations. Usually, total concentration of ice crystals are underestimated except for P3-1ICE at -30◦C and -45◦C. In the prediction, mixed-phased at -10◦C processes lead to an overestimation of LWC.

Response: We would like to express our acknowledgement for your efforts and constructive comments. Our point-by-point responses are given below. For convenience, the reviewers' comments are in **black** fonts, and our point-by-point responses are in **blue**.

Major Comments:

In this study, it is difficult to see the link between HIWC and their associated microphysical processes. Main of the figures and conclusion are built on figures that includes IWC>0.1g/m3, while the authors give the definition of HIWC such IWC>1g/m3. The comparison of size distribution is made for average PSD and for IWC > 0.1g/m3, which is a large spread. This study should at least separate the results with on one side figures and comment for 0.1g/m3<IWC<1g/m3 and in the other side for IWC>1g/m3.

Response: In this study, Figures 8-11 shows the results using TWC > 0.1 g/m^3 to evaluate the simulations, whereas Figures 12-14 show the results using TWC > 1 g/m^3 at −10 ºC to investigate microphysical processes associated with HIWC regions. We did not show PSDs for TWC > 1 g/m^3 at −45 and −30 ºC because the sample size is too small to be statistically significant. However, because there are enough samples at −10 ºC, we do separate the PSDs for 0.1 g/m^3 < TWC < 1 g/m^3 (Fig. R3-1) and TWC > 1 g/m^3 (Fig. R3-2). From the comparison of PSDs, the main conclusion for 0.1 g/m^3 < TWC < 1 g/m^3 and TWC > 1 g/m^3 (Figs. R3-1 and R3-2) are consistent with that for TWC > 0.1 g/m^3 (Fig. 10), namely that all of the simulations miss the peak of the observed PSD near D_{max} of 0.3 mm. According to your suggestion, we have added a brief description of these results in the revised manuscript, which reads "There are enough samples with TWC > 1 g m^{-3} at −10 ◦C to examine PSDs for 0.1 g m^{-3} < TWC < 1 g m^{-3} and TWC > 1 g m^{-3} separately. Both results indicate that all of the simulations miss the peak of the observed PSD near D_{max} of 0.3 mm (Figs. S1 and S2), which is consistent with that for TWC > 0.1 g m^{-3} (Fig. 10)."

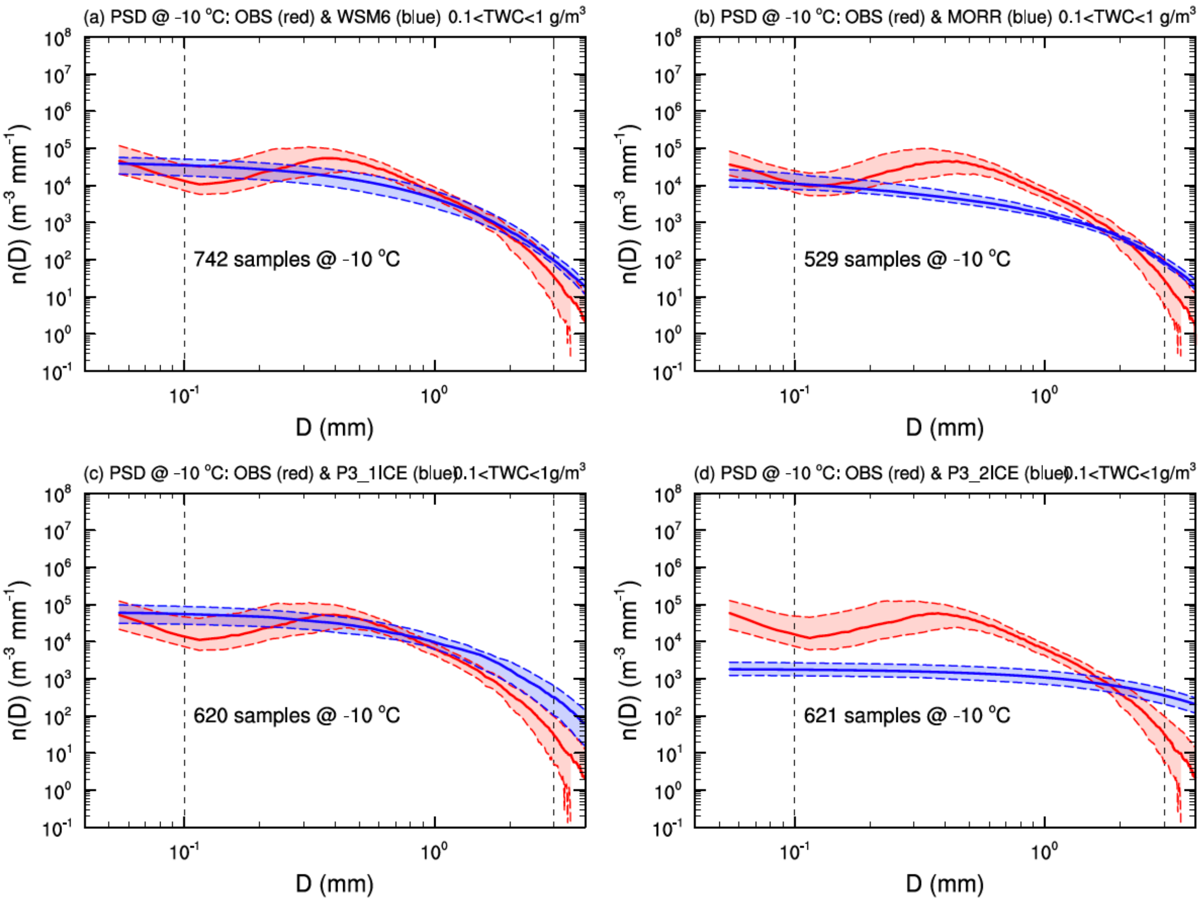


Fig. R3-1. Observed (red) and simulated (blue, a: WSM6, b: MORR, c: P3-1ICE, d: P3-2ICE) ice particle size distributions for 0.1 g/m^3 < TWC < 1 g/m^3 at the level of −10 ºC. The red and blue dashed lines indicate the 25th and 75th percentiles, and the red and blue solid lines represent the median.

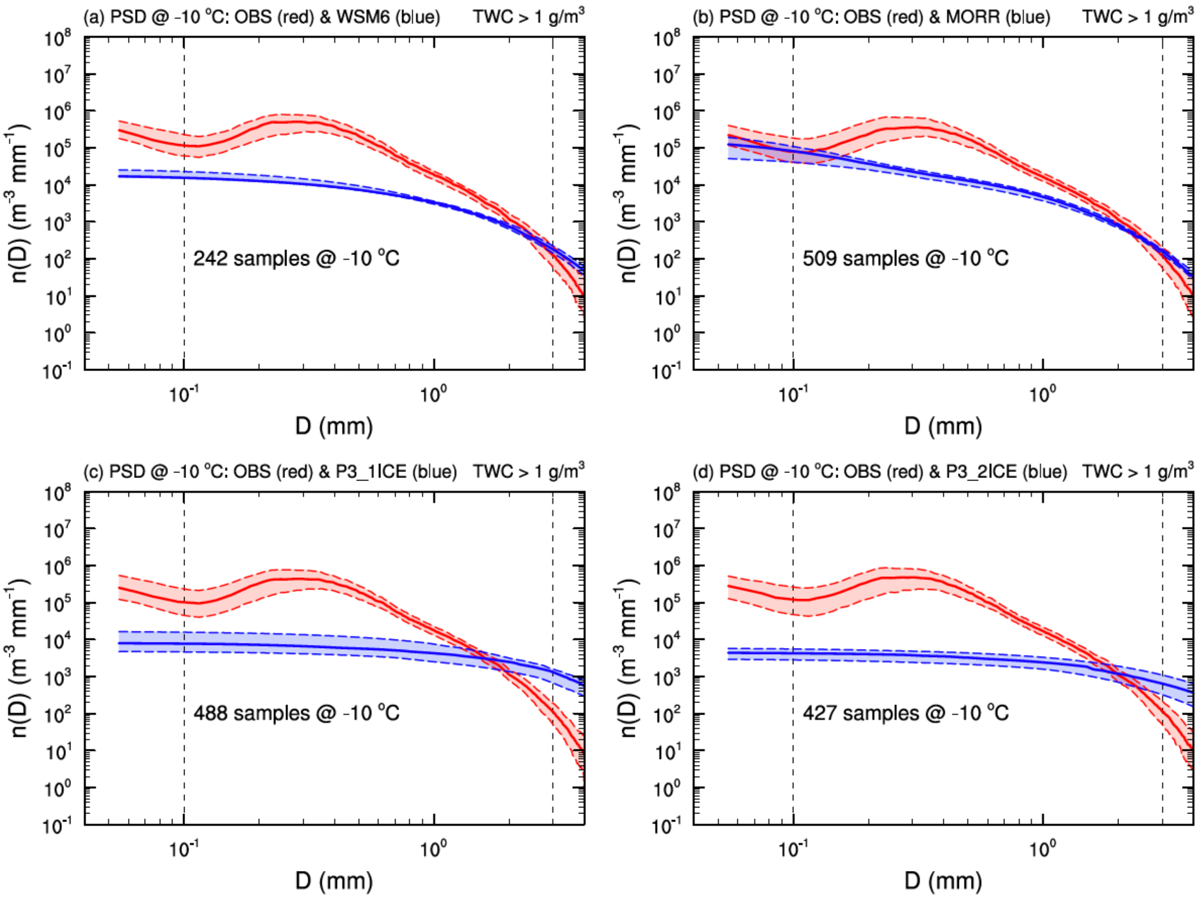


Fig. R3-2. Observed (red) and simulated (blue, a: WSM6, b: MORR, c: P3-1ICE, d: P3-2ICE) ice particle size distributions for TWC > 1 g/m^3 at the level of −10 ºC. The red and blue dashed lines indicate the 25th and 75th percentiles, and the red and blue solid lines represent the median.

In the introduction, the authors bring the attention on what HIWC at high altitudes are made with small ice crystals (~300-400µm) and that X band are not suited for detection of HIWC. But in the present study, the authors still use the X-band radar to evaluate the distribution of IWC in their simulations, while a look in the former publication that dedicated studies on HAIC-HIWC dataset cited in the introduction, give the information that a cloud radar was on board with the microphysical probes. Knowing the size of ice crystals that made mainly HIWC and the strong relationship demonstrated between Z, IWC and T, cloud radar is more suited to evaluate the distribution of IWC. Comparison of Z in X-band only allows to study aggregates densities and their distribution, thus aggregation process and precipitation. As Z at these frequencies (~10GHz) are more sensitive to the concentrations of large hydrometeors and not the total water content itself (see Drigeard et al., 2015).

Response: The main objective of this study is to evaluate the simulations using different bulk microphysics schemes with differently-parameterized PSDs of ice species using observations from radiosonde, satellite, radar, and in-situ microphysical probes. X-band radar reflectivity is mainly used to investigate reflectivity biases between observations and simulations, which is only one part of our evaluations. To avoid any confusion, we have clarified in the introduction that Wolde et al. (2016) found that "the ***pilot*** X-band weather radar on the Canadian National Research Council (NRC)'s Convair-580 aircraft did not have adequate sensitivity to detect HIWC regions when calibrated using the NRC X-band ***research*** radar." There are two different X-band radars on NRC Convair-580, the "***pilot*** X-band weather radar" and "X-band ***research*** radar". We have emphasized this difference by italicizing these words in the revised manuscript. We also note that Wolde et al. (2016) found that "For the Convair-580 Cayenne radar dataset, ***in cloud segments with HIWC exceeding 1 g m^{-3}, there is no significant difference between W and X band Ze values near the aircraft level***, which is again consistent with other publications and pilot reports where the MMD is less than 500 µm and the reflectivity less than 20 dBZ and within the Rayleigh scattering limits at both frequencies." The main differences between W and X band Ze values are in the melting layer dominated by aggregates and in rain regions below due to severe attenuation of the W band. Examining radar is not the focus of our study. Instead, we use radar reflectivity data from the "X-band ***research*** radar", which is suitable for HIWC (Wolde et al. 2016). We evaluated the IWC using in-situ data from microphysical probes on the NRC Convair-580 and SAFIRE Falcon 20, not using the radar retrievals which have larger uncertainty themselves and are often evaluated against in-situ data.

The comparison between observations and simulations is fair as long as X-band radar reflectivity forward operator is used for simulations.

Reference:

Wolde, M., C. Nguyen, A. Korolev, and M. Bastian, 2016: Characterization of the Pilot X-band radar responses to the HIWC environment during the Cayenne HAIC-HIWC 2015 Campaign. *8th AIAA Atmospheric and Space Environments Conference*, 8th AIAA Atmospheric and Space Environments Conference, Washington, D.C., American Institute of Aeronautics and Astronautics.

Overall, the conclusions are mainly known, overestimation of Z in C/X-bands, overestimation of LWC and then too much riming (see cited publications in the introduction; C and X bands differs from their radar constant and attenuation, but their response with regards to the hydrometeors are similar) and mainly underestimation of total concentrations due to a lack of SIP processes in convective clouds. However, the fact that P3-1ICE and P3-2ICE can produce high concentration of ice crystals is interesting. What SIP these two schemes take into account?

Response: While we agree that some previous studies have indicated reflectivity and ice size biases in simulations of tropical deep convections using different models, microphysics schemes and field studies to hypothesize that including SIP processes could improve previous simulations, few prior studies have comprehensively analyzed different microphysical schemes to assess the role of different microphysical processes in generating these biases. Our study is the first to test the success of the P3 scheme in simulating the HIWC phenomenon and to determine whether the P3 scheme is better able to predict HIWCs in a high-resolution numerical models. Collision/collection between ice categories and rime-splintering processes are taken into account in the P3 scheme, however these processes are activated only in P3-2ICE.

The authors mentioned in the introduction a companion paper more dedicated to the schemes P3. I suggest that the actual paper is withdrawn and will be re-submitted in the same time than its companion paper, including in their titles “part 1” and “part 2”.

Response: Although the work of second paper is related to this paper, they are separate articles with different motivations. A detailed study on P3 scheme is ongoing and will be submitted later this year. We feel that the conclusions in this study are significant enough to merit publication now and there is no need to delay until a subsequent study is ready for submission. Even many part 1 and part 2 papers are published at different times.

Minor comments:

Page 2, line 43-49: Does this IWC-Z-T relationship is in X or C band? How does it compares with the one of Protat et al., 2016 cited in line 35 of the same page? Same for the methodology from Nguyen et al. 2019, how does it compared with the one of Protat et al. 2016, mean bias, rms of both methods? Moreover, authors do not use X-band polarized parameters as it seems to be suggested by Nguyen et al. 2019 in order to study IWC, why?

Response: The IWC-Z-T relationship in Wolde et al. (2016) is for both X and W bands. Both Wolde et al. (2016) and Protat et al. (2016)'s IWC-Ze relationship followed a power-law fit with coefficients dependent on temperature, although they used different radar data from different aircraft at different temperature ranges at different locations. For the retrieval errors, in Protat et al. (2016): "Using a single IWC–Z95 relationship allows for the retrieval of IWC from radar reflectivities with less than 30% bias and 40%–70% rms difference. These errors can be reduced further, down to 10%–20% bias over the whole IWC range, using the temperature variability of this relationship." In Nguyen et al. (2019), "the use of the specific differential phase (Kdp) and differential reflectivity ratio (Zdr) significantly reduces errors in IWC retrieval over the conventional IWC–Z method", and for all data, the bias is -0.045 g m^{-3} and the rms is 0.52 g m^{-3} in their study. However, most of HIWC data points used in Nguyen et al. (2019) are measured at a narrow window of the temperature range (-10ºC ± 2.5ºC).

We did not use X-band polarized parameters for our study, because the X-band dual-polarization airborne radar is side-looking, as opposed to the X-band single-polarization airborne radars that are nadir and zenith pointing. The objective of this paper is to evaluate the vertical profiles of simulated radar reflectivity and not to retrieve IWC, so data from the nadir- and zenith-looking X-band airborne radars were used. Further, if we were to evaluate the simulated polarized parameters, we would need to develop forward operators for these polarized parameters, which are not mature currently and have large uncertainties especially for ice species. This is beyond the scope of the current paper.

Page 5-6, Section 3.2: It would help to add a description of main microphysical process and the SIP that are taken into account by each schemes. It will helps later, to discuss why P3 schemes have such high concentrations at high altitudes (Figure 11b)?

Response: We have added a description of the main microphysical processes associated with ice particles in each scheme to Section 3.2, which reads "In WSM6, main microphysical processes associated with ice particles include ice nucleation, deposition, sublimation, homogeneous and heterogeneous freezing, collection by liquid and other ice categories, autoconversion to other ice categories, melting, and sedimentation (Hong and Lim, 2006). MORR also considers rime-splintering process in addition to the same main microphysical processes as WSM6 (Morrison et al., 2009). Both P3-1ICE and P3-2ICE consider the main microphysical processes associated with ice particles including ice nucleation, deposition, sublimation, homogeneous and heterogeneous freezing, collection by liquid, self-collection, melting, and sedimentation, while collision and collection between ice categories and rime-splintering processes are considered only in P3-2ICE (Morrison and Milbrandt, 2015; Morrison and Milbrandt, 2016)." High ice number concentrations at high altitudes can be associated with homogeneous freezing of cloud droplets if they are lofted to higher levels (Fig. 14b).

Page 6, section 3.3: As commented in the major comment, evaluation of only Z in X-band can only help to study aggregation and precipitation processes. W-band is more suited for the topic of this study.

Response: Please see the response to the major comment.

Page 7, lines 187-215: Is there a relation between cloud top temperature and the IR brightness temperature? Does it means that the 4 schemes have a good predictions of the height of the convective clouds? What are cloud top heights that the radar estimate with regards to the co-located Tb?

Response: Brightness temperature is "a descriptive measure of radiation in terms of the temperature of a hypothetical blackbody emitting an identical amount of radiation at the same wavelength" (https://glossary.ametsoc.org/wiki/Brightness_temperature). The relationship between the brightness temperature (BT) and real temperature (T) is BT = εT, where ε is emissivity of the surface. Since $0 \leq \varepsilon \leq 1$, the real temperature will be greater than or equal to the brightness temperature.

Brightness temperature can indirectly indicate cloud top height. In general, the higher the cloud, the lower its brightness temperature. From the comparison between observed and simulated brightness temperature (Fig. 4), simulations with Morrison and P3 schemes have relatively larger areas than observations at the same brightness temperature, indicating these simulations slightly overestimate the cloud top heights. These results are consistent with the radar reflectivity vertical distribution (Fig. 7).

Page 8, lines 216-253: Does the same range of IWC in observations and predictions give the same range of Z, for the same T. It may need to be completed with cloud radar observations and comparison?

Response: In this study, we used in-situ IWC from microphysical probes installed on the aircraft to evaluate the simulations and did not evaluate IWC through radar retrievals. In fact, such retrievals have larger uncertainty dependent on retrieval method and are often evaluated by in-situ data. We used a forward operator to attain simulated radar reflectivity for evaluation of the model results.

Page 9, section 4.2: from figure 11a, it seems that range of IWC are not exactly the same with similar distributions as function of microphysical schemes and observations. The radar comparison shows a large variability of Z and a distribution of Z different between observation and prediction. Does it make sense to compare PSD over all IWC? I suggest a comparison of PSD as function of IWC range? The overestimation of concentrations of large hydrometeors can explain the figure 7. The four scheme could predict similar density as in observation, but the fact that the prediction of concentration of large hydrometeors are larger than in the observations is enough to understand the overestimation of Z ? What is the impact of IWC in the overestimation of Z? Note that P3-1ICE at -30◦C and -45◦C predict similar concentrations, or more exactly share a similar range of concentration with observations for large hydrometeors and that predicted Z are closer with the one observed (figure 7) than the other predicted Z.

Response: We select simulated TWC/PSD samples based on aircraft locations and brightness temperature fields and thus are comparing IWCs in similar locations. Of course there are still some biases between observed and simulated TWC. However, we think the TWC biases would not influence statistic PSD biases. Comparing the PSDs shown in Figs. 8−10 to the PSDs normalized by TWC at −45, −30 and −10 ºC (Figs. R3-3−5), we can see that the qualitative results are consistent with each other. We have added a brief description of these results in our revised manuscript, which reads "These qualitative results do not change through examining the PSDs normalized by TWC at −45, −30, and −10 ◦C (not shown)".

We should clarify that observed radar reflectivity profiles are sampled along the flight track of NRC Convair 580, whose horizontal locations are different from those of observed TWC/PSD samples from SAFIRE Falcon 20 (Fig. 1). Part of observed TWC/PSD samples at −10 ºC and all observed TWC/PSD samples at −45 and −30 ºC are from SAFIRE Falcon 20 (Fig. 1), implying that observed TWC/PSD at −45 and −30 ºC may be not fully consistent with observed radar reflectivity in the upper levels in this study. Comparing CFADs of observed and simulated radar reflectivity profiles indicates statistical biases in reflectivity. Comparing different percentiles of PSDs at different temperature levels reveals statistical PSD biases, which help us understand possible causes of reflectivity biases. From equation (A11), it implies that mass concentrates at the large end of PSD resulting in the overestimate of reflectivity. With a gamma or exponential distribution and constant m-D relation assumptions, reflectivity is proportional to IWC^2/N_t referring to equation (A16). Therefore, we can hypothesize that the overestimate of reflectivity results from IWC concentrating at the large end of PSD. The simulated reflectivities in upper levels in P3-1ICE are closer to the observations, which may be associated with similar simulated concentrations to observations at −45 and −30 ºC. We have added a brief description of these points in our revised manuscript, which reads "It should be noted that observed radar reflectivity profiles were sampled along the flight track of NRC Convair 580, whose horizontal locations differ from those of the flight track of the SAFIRE Falcon 20 (Fig. 1a). Part of observed TWC/PSD samples at −10 ◦C and all observed TWC/PSD samples at −45 and −30 ◦C are from SAFIRE Falcon 20 (Fig. 1b), implying that observed TWC/PSD at −45 and −30 ◦C may be not fully consistent with observed radar reflectivity in the upper levels in this study."

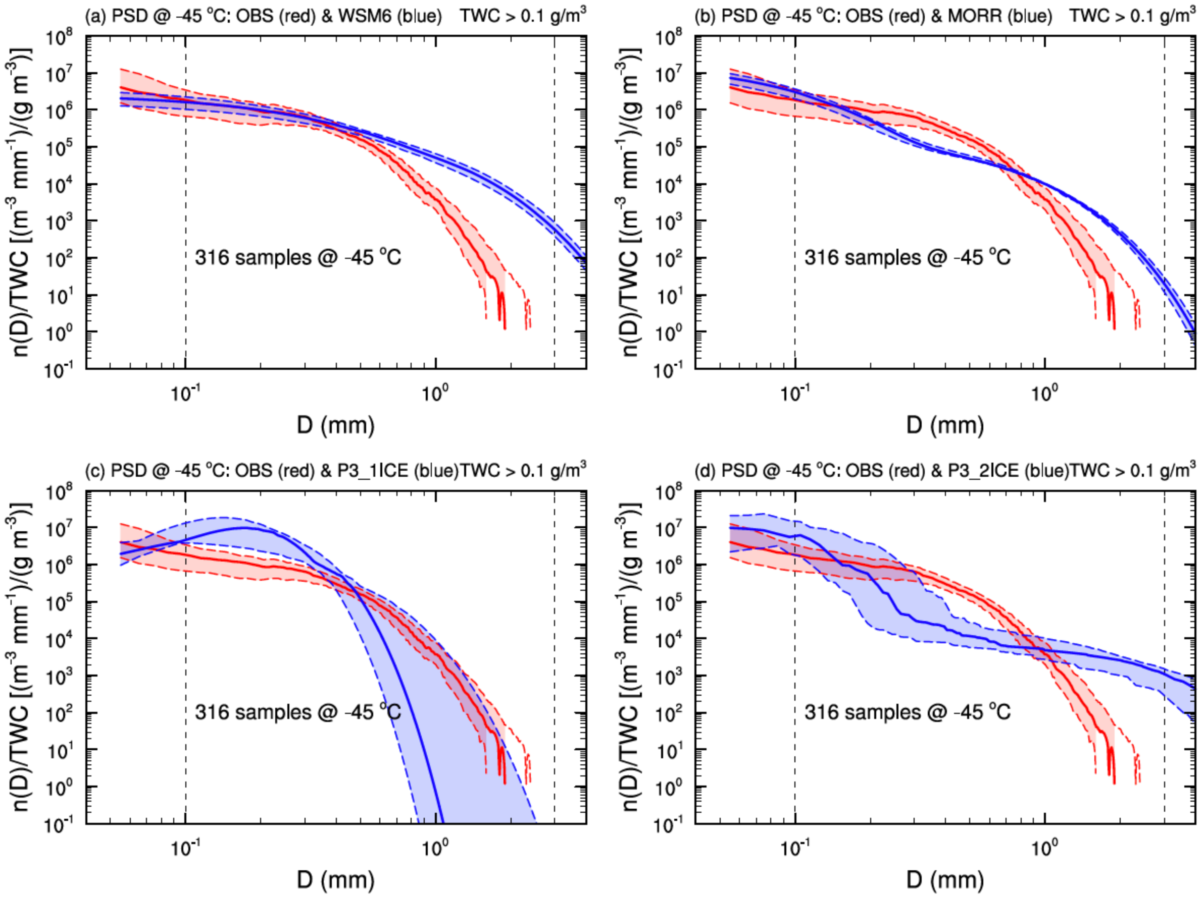


Fig. R3-3. Observed (red) and simulated (blue, a: WSM6, b: MORR, c: P3-1ICE, d: P3-2ICE) ice particle size distributions normalized by TWC at the level of −45 ºC. The red and blue dashed lines indicate the 25th and 75th percentiles, and the red and blue solid lines represent the median.

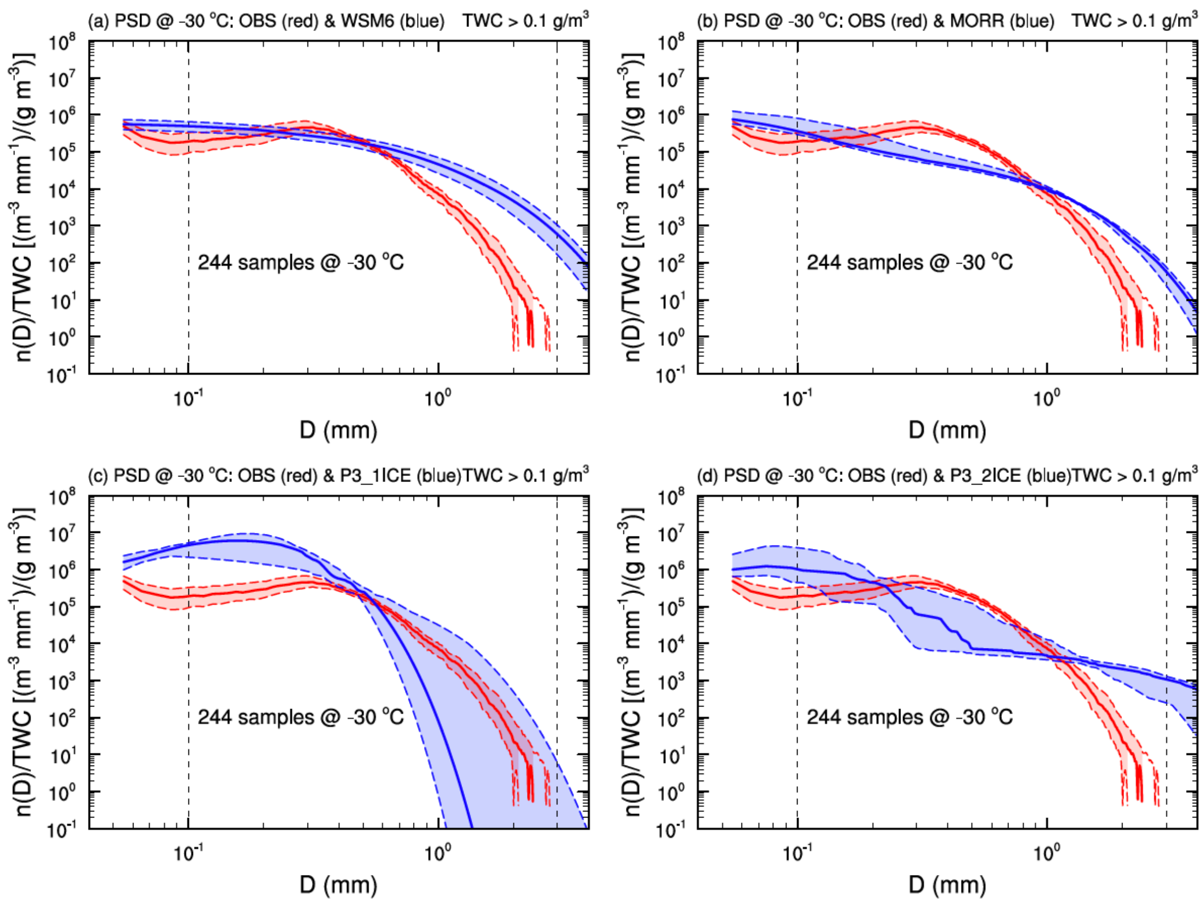


Fig. R3-4. Observed (red) and simulated (blue, a: WSM6, b: MORR, c: P3-1ICE, d: P3-2ICE) ice particle size distributions normalized by TWC at the level of −30 ºC. The red and blue dashed lines indicate the 25th and 75th percentiles, and the red and blue solid lines represent the median.

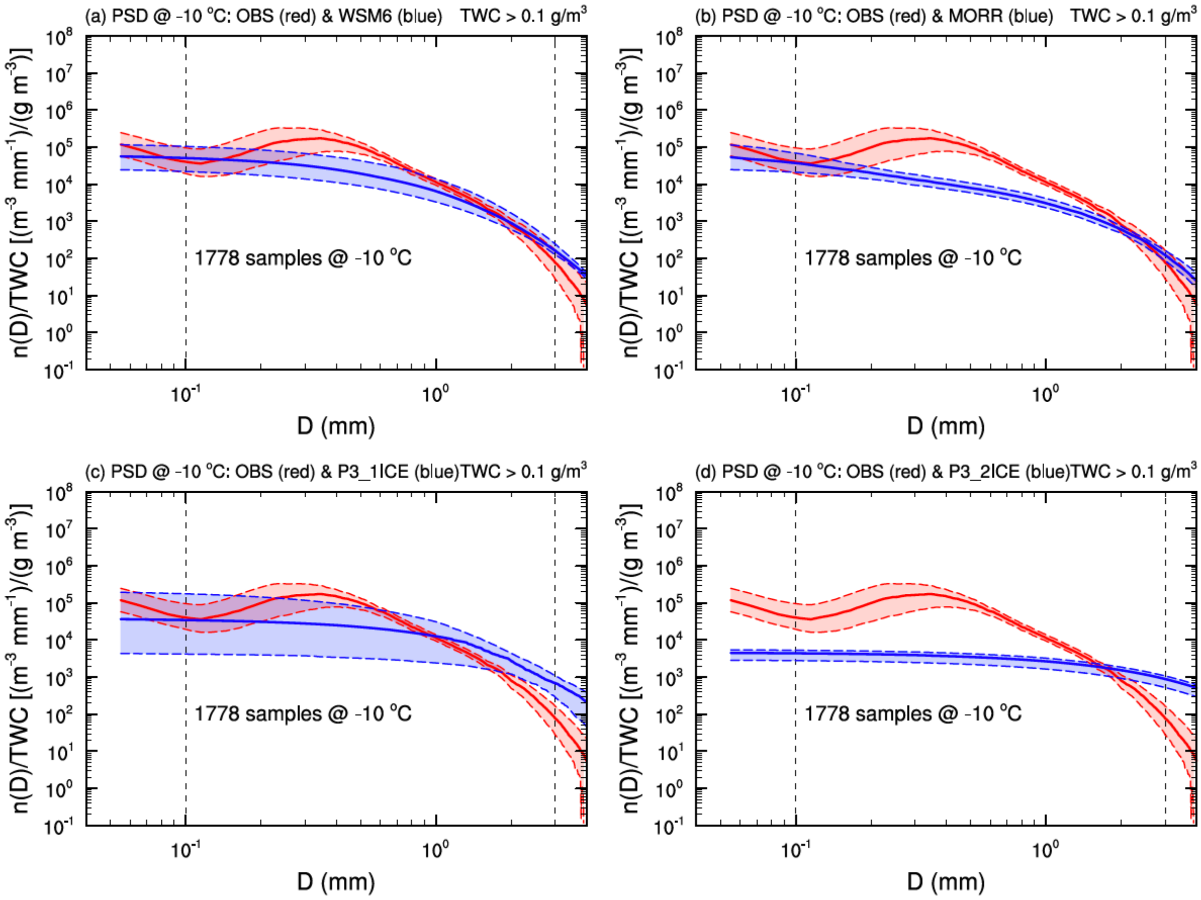


Fig. R3-5. Observed (red) and simulated (blue, a: WSM6, b: MORR, c: P3-1ICE, d: P3-2ICE) ice particle size distributions normalized by TWC at the level of −10 ºC. The red and blue dashed lines indicate the 25th and 75th percentiles, and the red and blue solid lines represent the median.

Page 14, line 417: Is this a new results?

Response: This paper mainly evaluates the simulations with different microphysics schemes. Although the summary #1 is not a new result compared to previous studies, it provides an evaluation of simulated dynamic and thermodynamic fields, which gives the context and foundation for the evaluation of other fields. Therefore, we prefer to keep this point in the summary.

Page 16, equation A2 + Figure 11b: Does the total concentration presented in Figure 11 from WRF prediction are calculated with this equation too ? The comparison would not be fair if it is.

Response: We derived the intercept (N_0), shape (μ) and slope (λ) factors of equation (A1) based on mixing ratio, total number concentration and PSD assumption of each scheme first, and then integrated equation (A1) for 0.1 mm < D < 3 mm. And the observed number concentration is also integrated only for 0.1 mm < D < 3 mm. Therefore, the comparison between observed and simulated $N_{0.1\text{-}3mm}$ is fair.