Review of the article titled “Coincident In-situ and Triple-Frequency Radar Airborne Observations in the Arctic” by Cuong M. Nguyen et al.

This article shows promising results of triple-frequency radar observations from the Radar Snow Experiment (RadSnowExp). Part of the uniqueness of this article is that both in situ and remotely sensed observations were collected by the same aircraft. Some complications arise when combining these datasets for their analysis, but the authors carefully and thoroughly describe the methodology used for volume matching and range calibration.

The authors studied the relationship between in situ sampled cloud microphysical properties and radar triple-frequency signals particularly, they centered their analysis on the relationships between median volume diameters, effective bulk density and dual frequency ratios. Lastly, the authors suggest a path forward with the possibility for quantitative retrieval of particle size using measured DFR but more in-depth analysis is needed to reach that step.

This article is generally well written, it is interesting, and it will be of interest to the scientific community. Nevertheless, I have a few concerns I suggest be addressed before this article is published. Below are some general and specific comments.

We thank the referee very much for the encouraging remarks. As some of the results and the dataset presented in this paper are new and unique, the need for clarity and attention to detail is appreciated. As such, we address below the concerns and comments expressed by this reviewer, to the extent we feel feasible.

General Comments:

The authors analyzed one flight (22 November) and divide it into 3 different segments (1948-2000 UTC, 2005-2028 UTC and 2121-2135 UTC) giving a total number of 49 minutes of DFR observations. Please consider this to put into context the overall findings that derive from this dataset that are stated in this article.

We thank the referee for this comment. We have added this information to the revised manuscript.

Figures need significant improvement, both in the actual quality of the figure (suggest improving ppi) and the legibility of axis labels and legends. For figure with more than one panel the different panels should be labeled. This was done for some of the figures but not in all. I suggest a uniform way to address multipanel figures and their caption. Finally, figure captions need to carefully state what is plotted in each panel/figure.

I’m not sure figures 1 and 6 are necessary if the article centers around the Nov 22 flight (in case fig 1 is kept, I made I suggestion below to have all the flights drawn or the domain where all the flights took place). Similarly, I don’t think figure 5 is necessary either, no extra analysis is done of how the overlapping sizes were treated or any extra analysis that would make this plot needed in the article, just stating in the text the size ranges each instrument measures should be sufficient.

We completely agree with the referee on this comment. We have done an intensive modification on the figures. The font size on figure labels and legends are increased. Panels on Fig. 13, 16, and 19 are rearranged and labeled for better readability and presentation. Additional information has been added to the figure captions. We have also removed Fig 1, 5, and 6 as they are not necessary for the paper as suggested. With this major modification we believe the revised manuscript would better facilitate the reading.

Specific Comments:

Lines 14-16: Consider stating the amount of data that was used to reach the conclusions stated in the article and not the overall flight hours of RadSnowExp.

We have revised the sentence as suggested.

Line 18: I'm not entirely sure this article showed how to accurately derive the level of riming from the DFR plane, consider rewriting this sentence.

We have rewritten this sentence.

Line 26: Add the definition of GPM-DPR

We have added the definition of GPM-DPR

Lines 29-31: Reference needed.

We have added Haynes et al., 2009; Hiley et al., 2011; and Matrosov et al., 2008.

Line 34: Suggest modifying this sentence as "The GPM Core Observatory carries…" or similar.

We have modified the sentence as suggested.

Lines 39-41: Reference needed.

The reference has been added.

Lines 63-65: Figure 1 does not support this statement, a figure showing all the flight tracks, or the experiment domain would be more useful.

As it is pointed out by the referee and others, we have remove Fig. 1 because it is not necessary for the paper. We have revised the text to reflect the change.

Line 71: Consider replacing 'uniquely' by 'unique'

We have corrected this error.

Line 71: Figure 3 should be Figure 2, please reorder figures.

In the revised manuscript, we have removed Fig. 1, 5 and 6 and changed the text accordingly.

Line 110: Should be figure 3 when re-ordered.

This error has been fixed.

Line 110: consider replacing ‘points‘ by ‘retrievals’ or similar.

We have replaced it with “data points”.

Lines 149-154: The second bullet point is not clear. What do you mean by ‘three radar data’? How was it mapped into a common range axis? How ‘reasonable’ is the homogeneity assumption? Was any sensitivity analysis done to evaluate it? A schematic of all the smoothing methods could be valuable to assess the actual volumes that are being compared.

We thank the referee for this comment. We have corrected the phrase “three radar data” to “data from the three radars” and also revised the text in this bullet for clarity. The common range axis has origin at the aircraft location and a 35 m grid. The data mapping is done using a standard interpolation method. Because the common range spacing (35 m) is greater than the sampling resolutions of the three radars the standard interpolation method works well. We found that the assumption of homogeneity of clouds is only needed within the largest radar volume and within 50 m around the aircraft. This assumption may not be met at large range and/or near the boundary of cloud/precipitation. In our case, the radar data used in the analysis are at 245 m from the aircraft. At this range, the radar volumes are small (less than 20 m in antenna beam width for a 4.2 deg beam) so that this assumption could be made.

We were aware that beam matching is a critical aspect for triple frequency radar analysis and considered the best approach to process the data. The impact of mismatch beams/non-uniform beam filling for the triple frequency analysis need to be addressed but it is beyond the scope of this paper. In the revision, we have removed the word “reasonable” to avoid the confusion.

Lines 175-179: I consider this to be an important factor in the data analyzed here that could grant the inclusion of a figure showing these values to support this statement.

We completely agree with the referee on this point. In the revised manuscript, we have added the scatter plots of cross-calibrated W, Ka and X reflectivities at 245 m from the nadir antennas in the 22 Nov flight for the region of median size < 300 um.

Line 180: Figures should be numbered in sequential order they are referenced in the text. Consider reorganize the figures or the text.

We have modified the text to address the figure order issue.

Line 181: Were there more than 1 flight on 22 Nov? How much data does this represent? i.e.: How many data points were used to reach this conclusion?

There was only on flight on 22 Nov. We used 510 data points where the aircraft sampled regions of small ice crystal (median size <300 um) to verify the cross-calibration between the frequencies. In the revised manuscript, we have included scatter plots of the reflectivities in the regions of small ice particles to illustrate the conclusion.

Lines 193-194: Consider adding sizes to contextualize 'small cloud droplets' and 'large precipitation hydrometeors'.

We have added sizes as suggested.

Line 197: Consider replacing 'or' by 'and'.

We have revised the text as suggested.

Line 206: Is the difference between all the 9 groups and the subset of ice habits only the inclusion or not of the Drops and Artifacts category? How are small particles treated?

The subset of ice only includes pristine, dendrites, rimed dendrites, rimed particles, aggregates and other ice particles. Small particles are not in the subset of ice habits.

Lines 212-213: It would be good to state the uncertainty values that are within the range presented by Baumgardner et al. (2017).

We computed the uncertainty values in sizing and concentrations from a wind tunnel calibration dataset. However, the uncertainty values in flight environment could be different.

Lines 219-224: This paragraph is confusing, if the wiring had little effect on the estimated water content, then what was the factor that made the accuracy drop from 0.002 g/m3 to 0.05 g/m3? Was this value taken as a constant value regardless of the size of the hydrometeors sampled? How was the estimation of the accuracy of the Nevzorov probe done?

We have revised the paragraph for clarity.

First, we would like to highlight that 0.002 g/m3 is the sensitivity reported by (Abel et al., 2014) for this commercial sensor, while 0.05 g/m3 is our estimated accuracy due to uncertainty in the dry term calculation. It is reasonable to expect that in the presented airborne measurement, LWC and IWC may be underestimated, especially in segments with higher MVDs. Unlike in wind tunnel studies, the quantification of this bias is challenging due to the complexity of verification of IWC measurements in naturally occurring ice particle population. In other flight campaign (e.g. Faber et al. 2018), similar uncertainty values were found. We included this clarification in the text.

Line 226: If the minimum of 50 µm was used as lower bound then FCDP was not used in this analysis?

The threshold 50 µm is only used in the calculation of effective bulk density. For MVD, number of concentration (Nt) and other qualitative assessments, we use the full spectrum of the PSD.

Line 229: Consider adding "(The definition of) several bulk…"

We thank the referee for this comment. We have revised the text as suggested.

Line 232: Add year to Heymsfield et al.

We have added the year to this reference.

Lines 242-248: I think this paragraph and figure 6 could be removed from the article. Just adding a sentence that states that 22Nov will be analyzed and why should be sufficient.

We agree with the referee. In the revised manuscript, we have removed Fig. 6 and amended the text accordingly.

Line 258: Consider adding "In (current or past) literature…" and adding several references of this literature.

We have added some references to this sentence in the revision.

Lines 266-267: Convair average ground speed is 100m/s, this makes the volume for the in situ sampling of 200-500 m. How is this mismatch between the radar and in situ volumes handled?

The in situ probe sampling volume (Baumgardner et al., 2017) is much smaller compared to the radar volume. We believe there is no processing technique to match volumes from the in situ probes and the radars. In our processing, the radar data is decimated to match with the temporal resolution of the in situ data.

Lines 271-272: I suggest analyzing if reflectivity shows a different signal comparing Nadir vs Zenith samples and then compare DFR.

In this flight, the aircraft stayed in inhomogeneous cloud layers most of the time. Figure 11 (pdf of the X-band reflectivity at the nearest range above and below the aircraft) shows the difference between the nadir and zenith data with higher reflectivities typically occurring below the aircraft. This is the reason we decided just to compare DFR directly.

Section 3.3.1: This paragraph and associated plot is confusing, the almost 500 m difference between the observations at nadir and zenith show that the vertical structure of the cloud has a large impact on DFR. This is particularly clear when comparing DFR X/Ka, where the Zenith ratio has an almost constant value regardless of the values measured at Nadir. Is there a reason for this? As mentioned before, joint distributions of reflectivity would be better to analyze this 500 m effect.

Unfortunately, as we have discussed in section 3.1.2, the closest useable triple frequency radar data are 245m below and above the aircraft. Hence, there is almost 500m between the DFR observations. The example (Fig. 7) shows that in mixed phased clouds the DFR could vary in a wide range in a short distance (<500 m). In this section, we want to emphasize that due to possibly large variability in DFR observations, matching in situ and radar data is critical. In the OLYMPEX airborne campaign (Chase et al., 2018) when they flew two aircrafts, the radar collected within 10 minute temporally and 1 km spatially of the in situ are regarded as collocated. Since we wanted to focus on the DFR variability, we opted to show the joint distributions of DFRs instead of reflectivies.

Line 288: Add ‘s’ to Gans.

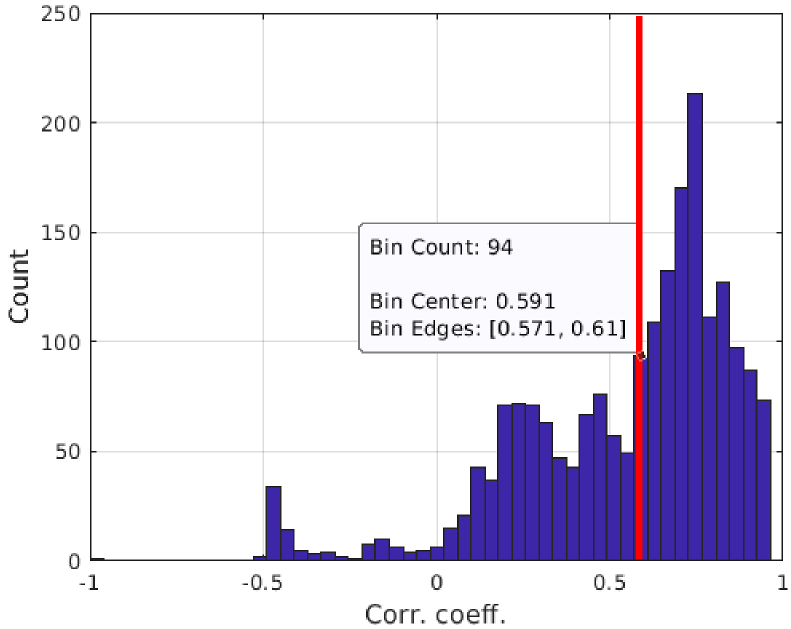
The typo has been corrected.

Line 294: Figures should be numbered in sequential order.

We have changed the text to address this issue.

Lines 292-294: I’d consider regions where both the Nadir and the Zenith correlation coefficient are good. This will hint at a more homogeneous cloud in the vertical and thus a more reliable comparison between what is sample in situ and at a 245 m difference in height. For example, the 4th region marked in figure 10 shows a big difference between Nadir and Zenith (based on correlation coefficient) this is most probably hinting that the part of the cloud sampled has a notable vertical structure so, I’m not sure it is a good case to analyze DFR because this gives an extra reason for the differences between the DFRs.

We agree with the referee that it would be best to select regions where both the nadir and the zenith correlation coefficient are good. However, because in most part of this flight we sampled inhomogeneous clouds, this condition is hard to be met. In the first and second flight segments, the nadir correlation coefficient is good (≥0.6) and the zenith correlation coefficient is also good for most part of the segments. In the last segment the nadir correlation coefficient is high whist the zenith correlation coefficient is greater than 0.5 most of the time so we still believe it is a good case for DFR analysis. We have attached a figure showing the histogram of the nadir correlation coefficient. The histogram show a main mode with correlation coefficient greater than 0.6. Hence, we selected a threshold of 0.6 for the decision of a good match between the in situ measurements and the radar data.



Lines 296-305: How long was this flight? How many samples were analyzed?

The flight was 3.5 hour and we analyzed 49 minutes (or 588 samples) of triple-frequency observations in the three study cases. We have added this information the manuscript.

Line 333: Replace ‘similarity measurements‘ by ‘correlation coefficient filter’ or cc threshold or similar.

We have revised the text as suggested.

Line 339: How where the different sections within segment 1, 2 and 4 defined? Was this breakdown into sections defined by the aircraft sampling pattern? Because figure 13 shows that there are different processes occurring in these different sections, for example, section D shows clearly different behaviors in DFR and CPI particle fraction near the beginning of the section when compared to the end of the section.

The decision for the breakdown into sections within the segments is based on the differences in the cloud processes, the observed DFRs, and the bulk measurements which are all related. We found that by naming different sections, it is easier to describe the cases and to present the results. It does not depend on the aircraft sampling pattern.

Lines 340-346: It’d be best if this description of the first segment has figure 13 as reference, it’d help contextualize the differences in the different segments.

We agree with the referee on this point. In the revised manuscript, we have combined figure 12 and 13 and amended the text accordingly.

Lines 345-346: The bimodality in the PSD distributions it difficult to see for all sections, especially the referred maxima at 1 mm. Please clarify.

In the revision, we have changed the text to “The mass distributions are generally bi-modal with two ice modes around 30 μm and 1 mm”

Line 351: IWC should be TWC or is the legend in the figure incorrect?

It should be TWC. The error is now corrected.

Line 357: Why/How was the aircraft at 6 km height? I assumed the black line in Figure 10a is the aircraft path, so in sections b-d shouldn’t the aircraft be at around 2.4 km height? It would be extremely beneficial to have the different sections shown in figure 10a.

We thank the referee for pointing out this error. It should read “at about 2 km”. We agree with the referee having the different sections shown in figure 10a is helpful. We tried to implement this idea but the figure is getting too busy; especially when we also added modelling lines to those figures. We might try different idea to address this concern in the final version of the manuscript.

Line 364: Consider rewriting ‘remarkably mirrors’

We have changed it to “resembles”.

Line 365: Similar to the comment before, from Figure 10a after the first few minutes of the first segment (first few minutes of section a that the aircraft descended) the aircraft seems to be flying at a constant altitude of ~2.4 km, is this not the case? Aren’t sections A-D correspond to segment 1 that is shown by the first box in figure 10a?

We apologise for this confusion. The text was messed up from another section. It is correct that the aircraft stayed at a constant altitude (~2.4 km) during sections B-D. We have corrected this mistake.

Lines 372-373: This sentence that the fraction of dendrites and rimed particles drops to its lowest level in section D can be misleading, this is the case near the beginning of the section, but by the end of the section this is clearly not the case. Consider rephrasing to avoid confusion.

We have changed the text to “ … rimed particles drops to its lowest level at the first half of the section when the highest DFR X/Ka occurs” for clarify.

Lines 390-391: Would it be possible to add to figure 14 the line representing graupel particles using discrete dipole approximation? It could be helpful to add the relevant curves from figure 2.

We agree with the referee. The relevant curves from Fig. 2 have been added to Fig. 14, 17, and 20.

Line 405: Consider ordering figures in sequential order that are mentioned in the article (Figure 15 should come before Figure 16). Also, how where these different sections defined?

In the revised manuscript, we have merged Fig. 15 into Fig. 16. We believe it would help the readers follow the analysis easier.

Lines 409-412: From figure 16 top right panel it seems like the fraction of dendrites is higher in C than in B?

We thank the referee for pointing out this error. The sentence has been corrected "In section B, the fraction of rimed particles is the highest."

Line 419: MDV does not reach 6 mm in section A, please clarify.

We have amended the text to correct this error. It now reads "…and DFR X/Ka at times reached the same level as DFR Ka/W at around 8 dB when MVD exceeds 4 mm."

Lines 420-421: ZDR was not mentioned before in the article, was there not any ZDR signature in the previous segment?

We thank the referee for this comment. We have added a sentence on the ZDR observation for the first segment.

Line 421-423: Consider rewriting this sentence, variables do not mimic other variables. Also, MDV is not > 8 mm for all the times that DFR is ~ 10 dB, this occurs just for the second maxima in DFR in section B.

We have changed the text and also corrected for the value of MVD " … and MVD is greater than 6 mm".

Line 423: consider adding a time series of ZDR to figure 16.

The time series of ZDR was already plotted in the second panel of the CPI plots in Fig. 16.

Lines 426-427: What do you mean by 'fluctuations in the DFRs'? Also, what datapoints correspond to section A, B or C is not clear from the plot, consider making the markers edge a different color linked with each section. Also, I suggest adjusting the limits of the plot to better fit the data plotted this will make the differences in the markers size and colors clearer.

We have rephrased the text 'fluctuations in the DFRs' in the revision. Regarding the scatter plot, we would like to keep the axis limits consistent between all the study cases and the modelling work (Fig. 2). In the revised manuscript, we have used different colors for the edge the dots to help identify sections A-C easier.

Lines 429-435: These few sentences are confusing consider rewriting. The hook feature is present in data from section C not B. Also, remove parenthesis for the references inside the larger parenthesis, like in the Petty and Huang, 2010 reference.

We have rewritten the paragraph for clarity. Data from all the section (A-C) present a hook feature and the turning point occurs at section. We have also corrected for the redundant parenthesises in the references.

Line 444: Why was this segment chose to be analyzed in depth? Please consider the previous comment regarding the difference between observations at Nadir and Zenith with respect to the vertical structure of the cloud and how different processes can be playing a role at different heights of the clouds that could give difference in DFRs that are not exclusively related to the factors analyzed here. This variability in height can clearly be seen in figure 10a where different microphysics might be acting between the lowest trusted range sampled by the radar and the in situ observations sampled 245 m below.

In this segment, we sampled a region of high concentration of supercooled drop with heavy ice accretion and then a region of milimetric rimed particles and large aggregates. The shape of the hook feature in Fig 19 is different from the previous study cases. In section C of this segment, DFR Ka/W goes up to 10-12 dB at MVD ~ 6 mm similar to the previous segment but DFR X/Ka is much lower at 5-7 dB. This is a feature which, we believe, is worth to be analyzed in depth. We agree with the referee that in this segment, there is difference in the nadir and zenith observations but as mentioned in a response before, we decided to use this segment because the nadir data is identified as good match based on the correlation.

Lines 484-485: This sentence is misleading as a summary and discussion part of this article where 1 day was analyzed and 3 segments of that flight that resulted in 49 minutes of DFR observations.

We have added “The whole RadSnowExp dataset …” to avoid the confusion with the flight we analyzed in this paper. The next sentence should make it clear to the readers “The potential of this dataset is illustrated here using one flight data during an Arctic storm that covers a wide range of snow habits from …”

Lines 488-490: This sentence is confusing, please consider rewriting it.

We have rephrased this sentence.

Lines 496-499: Add ‘for the flight we analyzed’ or similar phrase here as the results described could be case dependent.

We have added the text as suggested.

Lines 503-504: Figure 22 should be figure 21?

Yes. Correction has been made.

Line 503: I'm not sure I understand how figure 22 (or 21) is a first attempt for quantitative retrieval of particle size using measured DFR Ka/W and DFR X/Ka. There are a lot mean MVD and density values that are linked with different DFRs values. This looks more like a qualitative analysis.

We have revised the sentence for clarity. It now reads "Reversely, the strong connection between the particle size and the triple-frequency radar signature suggests that the data could be directly used to produce look-up-tables for mapping measurements in the (DFR Ka/W, DFR X/Ka) space into microphysical properties like median volume diameter and effective bulk density with associated uncertainties.". We also added standard deviation plot of the parameters to Fig. 21.

Line 505: Figure 21b shows that equivalent density increases with decreasing both DFR

Actually that sentence is not needed for the discussion in this paragraph. In the revision, we have removed it. The discussion of the ρ_e rotation feature is better illustrated in study case 3.

Line 510: Remove extra '.'

This error has been corrected.

Lines 515-517: Change 'demonstrated' by 'shown', otherwise, this is too strong of a statement based in the case analyzed in this article.

Correction has been made.

Lines 523-524: This sentence is not clear, consider rewriting it for clarity.

The sentence is not necessary for the discussion and has been removed.

Fig. 1: please improve figure 1a, it is very difficult to visualize the location of the flights, what is the color in the track representing? Not sure how convoluted a figure showing the path of all the flights done in RadSnowExp, but it'll sure be useful to see such figure.

In the revised manuscript, we have removed Fig. 1, 5 and 6 because they are not necessary for the paper.

Fig. 4: Consider center this figure around the heights that are referenced in the article. For example, if the ground is not referenced why extend the y-axis all the way 5 km?

We have modified this figure to include a "zoom in" section at the close range region. The ground is shown to illustrate the three radar profiles are well aligned (section 3.1.1).

Fig. 7: Please add in the figure caption what is shown in the figure (what is the dashed line and the error? bars).

The dashed line is the simple 1:1 line and the black line presents the data means of the DFRs with error bars of one standard deviation. We have added this information to the caption of Fig. 7.

Fig. 8: Consider adding lat and lon to the plots and make the maps larger the synoptic map is not very useful in the context given for people not familiarize with the area the flights took place to easily link all 3 maps.

We have added the lat and lon lines to the Fig. 8b.

Fig. 9: What is the F9 legend? What do you mean by ‘the ground to air temperature’ ground temperature is ~ -23C and why is this important information to have?

Fig. 9 has been revised. The temperature is now plotted as a function of altitude. Proper legend has been included.

Fig. 10: Improve figure caption to add what are the red, blue and black lines in panel (b) and also improve the legibility of the legend in panels b-d. Also, please consider not using similarity measurement and just use correlation coefficient.

We have increased the vertical size of the figure and the font size of figure 10 to improve the readability. The caption has been revised as suggested.

Fig. 12: Please correct what figure is referencing the five sections. Also, I don’t see a reference in the article that requires panel b to be part of the figure.

In the revision, we merged Fig, 12 and 13 into one figure to facilitate the reading. The Fig. 12b would be referred in line 346 of the original manuscript.

Fig. 13: Please label each panel of the figure for an easier read and to improve the connection between the text and each plot in the figure.

We have modified Fig. 13 significantly as suggested by the referee and other reviewers.

Fig 14: Consider modifying the colorscheme of the scatter plot to improve the readability of the figure (pinkish colors represents both low and high values).

The figure’s colormap and display scale have been changed for better readability.

Fig. 15b: I don’t see a mention to this panel in the article.

It is mentioned in line 429-431 in the manuscript. In the revise manuscript, we have combined Fig. 15 and Fig. 16.

Fig. 16: Similar to Figure 13 please label each panel.

We have added labels to all the panels.

Fig. 21: A joint distribution with number of samples would give reference to the mean MVD and density.

We have added number of samples used to compute the mean MVD and density.

Fig. 22: Add what the black line and vertical black lines are in the figure caption.

We have added the description of the black line to the caption of Fig. 22.