Responses to the reviewer #3

Overview:
In addition to being interesting, I think this study is well thought out and the manuscript is well written. I found reading it to be a pleasure. In particular, I would like to compliment the authors on the quality of the introduction. That said, I do have one major concern (general comment #1, below) and several suggestions that might improve the manuscript.

Recommendation: Publish after major revisions

I have labeled this as a major revision, out of concern that responding to general comment #1 may require regenerating most of the figures in the manuscript and I want to give the authors ample time to do this, if necessary. This should not be taken to mean this is a poor work. I think it is excellent.
Reply: We thank this reviewer for the encouraging, insightful and constructive comments, which really help improve the manuscript significantly.

Before addressing your comments/questions below, first we would like to provide a summary of the major revisions made to the manuscript:

- We revised significantly the part about the bimodal joint distribution between qc and Nc in section 4. In particular, we pointed out that it is most likely just a coincidence that each side of the "V" shape track sampled one mode of the bimodal distribution. The along/across wind difference between the two sides is unlikely to be the cause of the bimodality.
- Three new cases that are either non-precipitating or weakly precipitating were added to the paper and they have no overall impacts on the conclusions. The flight track and radar reflectivity plots for all the cases, except for July 18, 2017, are provided in the supplementary material.
- A small bug in our code was found and fixed. This bug affects the computation of the EF based on lognormal distributions. As a result, the E_q based on the lognormal PDF agrees very well with the observation-based E_q (new Figure 6a), and the E based on the bivariate lognormal distribution agrees well with the observation-based E (new Figure 6d). Because of this, the Figure 8 was removed from the paper.
- Most figures are revised/updated per request/suggestion of the reviewers.

After these revisions, we think the paper is much improved and more focused, although the general conclusions still hold.

General Comments:
1) Bimodality in the relationship between qc and N
My only major concern with the analysis is that the bimodality in the relationship between qc and Nc has a significant impact on the results. The manuscript documents that this bimodality is associated with systematically different values for N measured when the aircraft was flying either along and across the wind. It seems to me that it is extremely unlikely that this would happen by random chance (that the aircraft just happened to observe a cloud field with two distinction regions with differing N that happen to align with the flight tracks) given that the same offset is found between legs occurring at different times and different altitudes. I can't think

of any physical reason why this should occur. As such it seems to me that it is very likely that this is a measurement artifact. Accordingly:

Reply: Indeed, this is a very interesting and puzzling phenomenon/observation. Although we believe the bimodality is "real" (in other words, it is unlikely an instrument artifact), we think it is most likely just a coincidence that each mode is sampled by one side of the "V" shape track. We do NOT think that the difference in wind pattern (i.e., along wind vs. across wind) "causes" bimodality. Instead, the two sides of the "v"-shaped happen to sample different regions of the "virtual" grid box that have different qc-Nc relations. One can imagine that, if the airplane can sample every point in the "virtual" grid box, then there would be no along-wind vs. across-wind (or from a different perspective, all points can be either along wind or across wind). So, there is no causal relation between along/across-wind difference and the bimodality. They are both results of "v"-shaped sampling pattern. On the other hand, it is important to note that a bimodal sub-grid joint PDF between qc and Nc can very well exist in the nature, which could be a result of sub-grid variation of, for example, vertical draft velocity, precipitation and/or CCN.

In short, we argue that the bimodal joint PDF of qc and Nc is "real", but it is not a result of the along/across wind pattern difference. We will further elaborate on this point in the context of your questions/comments. We have also revised the manuscript accordingly.

(A) Discussion: Is it real?
Am I missing something simple? In general, I think the possible causes of this systematic shift need to be discussed.
Reply: As we explained above, we do believe that the observed bimodal joint PDF between qc and Nc is "real". Perhaps, we should not have emphasized the along-wind vs. across wind difference between the two sides of the v shape flight, as it is most likely just a coincidence.

In the revised the manuscript, we point it out that the along/across wind difference and the bimodal joint PDF are likely coincidence. We also pointed it out that some sub-grid variations, for example, vertical draft velocity, precipitation and/or CCN can lead to bimodal joint PDF between qc and Nc.

(B) Subdivide the horizontal leg data between along and cross wind directions. All of the analysis is based on analyzing full length of the horizontal legs, which if I understand correctly mean the outer scale is about 60 km. That is, the enhancement factors you present represent the enhancement in going from the ~10 m scale of the FCDP measurement to a nominal grid scale of ~60 km. This is a choice you made. There is (as far as I can see) nothing that prevents you from dividing your hlegs into separate along and cross wind legs (I presume each will be about 30 km) and instead examine the variability on this smaller scale ~30 km scale. Obviously doing this will increase the sampling uncertainty associated with each leg, but you will have twice the number of points in most of your figures and more critically it will significantly reduce the impact of any measurement bias in N associated with the along or cross wind measurements.
Reply: Yes, your understanding is correct. We are using the high-resolution (~10 m) FCDP measurements to investigate the sub-grid scale variations of qc and Nc for a nominal GCM grid of 60 km. We made the choice *not* to separate the two sides of the "v" shape flight because the

current generation GCMs usually have a typical resolution of ~100km. A ~60 km grid box would be more relevant to climate modeling in this regard.

Before we submitted the manuscript, we had already done the sensitivity study as your suggested, i.e., separating the two sides of “v” shape flight for the July 18, 2017 case. The new hlegs after the resampling are shown in Figure 1 below. We selected the new hlegs #6 to #13 (corresponding to the original hleg #5 to #8) and #16 to #21 (corresponding to the original hleg #10 to #12) for further analysis. As listed in Table 1, the new hlegs 6, 9, 10, 13, 16, 19, 20 are from the west side of the “v” shape flight (along-wind) and 7, 8, 11, 12, 17, 18, 21 are from the east side of the “v” shape flight (across-wind)

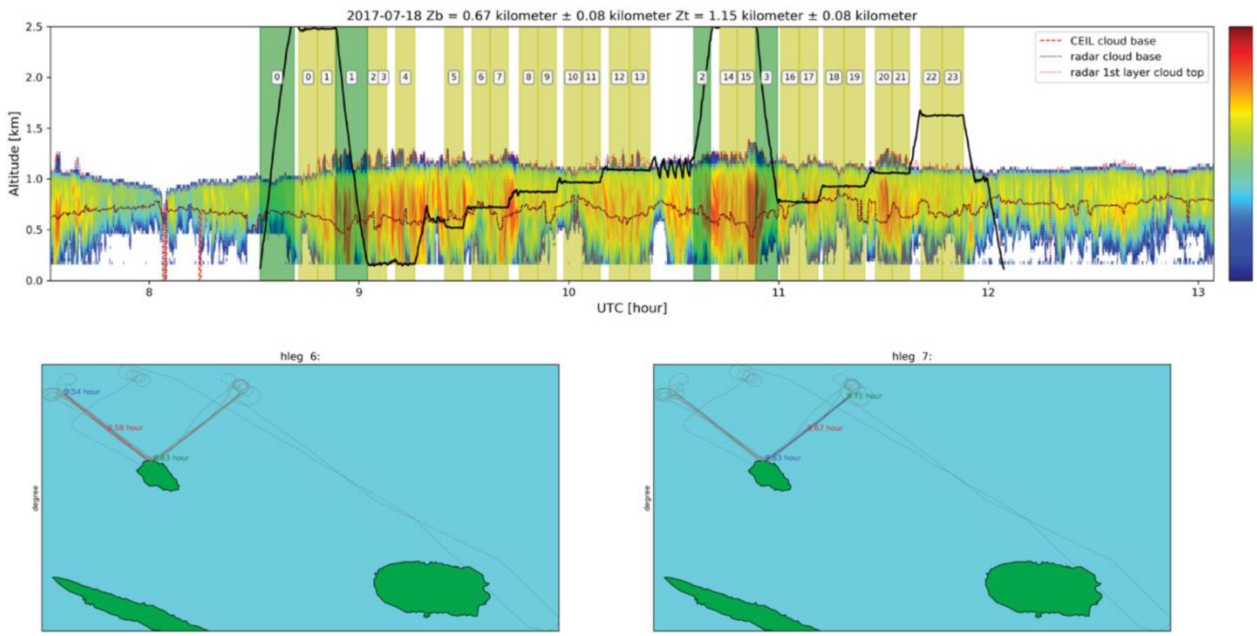

Figure 1 re-sampling of the hlegs for the July 18, 2017 case. Upper panel: The vertical flight track plotted on radar reflectivity measurements. Lower panel: an example to show that after the resampling, the new hleg #6 and #7 each corresponds one side of the “v” shape flight.

Table 1 The selected hlegs after the resampling

	West side of “v” shape (along wind)	East side of “v” shape (across wind)
New hlegs	6, 9, 10, 13, 16, 19, 20	7, 8, 11, 12, 17, 18, 21

After the resampling, the first thing we checked is whether the two modes of the bimodal joint PDF actually correspond to the two sides of the “v” shape flight. And indeed, it is the case.

Then, we repeated the same analysis as we did in Figure 4 and 6 of the original paper for the new hlegs, now with the two sides of the “v” shape flight separated. The key results are summarized in Figure 2. Evidently, both the west side hlegs (Figure 2 a-c) and east side hlegs (Figure 2 d-f) demonstrate the same vertical structures which are also consistent with those shown in Figure 4 and 6 of the paper: 1) the inverse relative variances ν_q and ν_N first increases from cloud base to cloud top and then decrees. 2) q_c and N_c become increasingly correlated from cloud base to cloud top. And 3) the total EF decreases from cloud base to cloud top.

These results clearly demonstrate that our conclusions are robust regardless whether or not we separate the "v" shape flight into two sides. We would like to hold onto the original results and plots based on the whole "v" shape flight in the manuscript for two reasons: 1) ~60 km spatial scale is more relevant to the current GCMs and 2) the bimodal joint PDF is possible and its implications for EF and thereby warm rain simulation should be discussed.

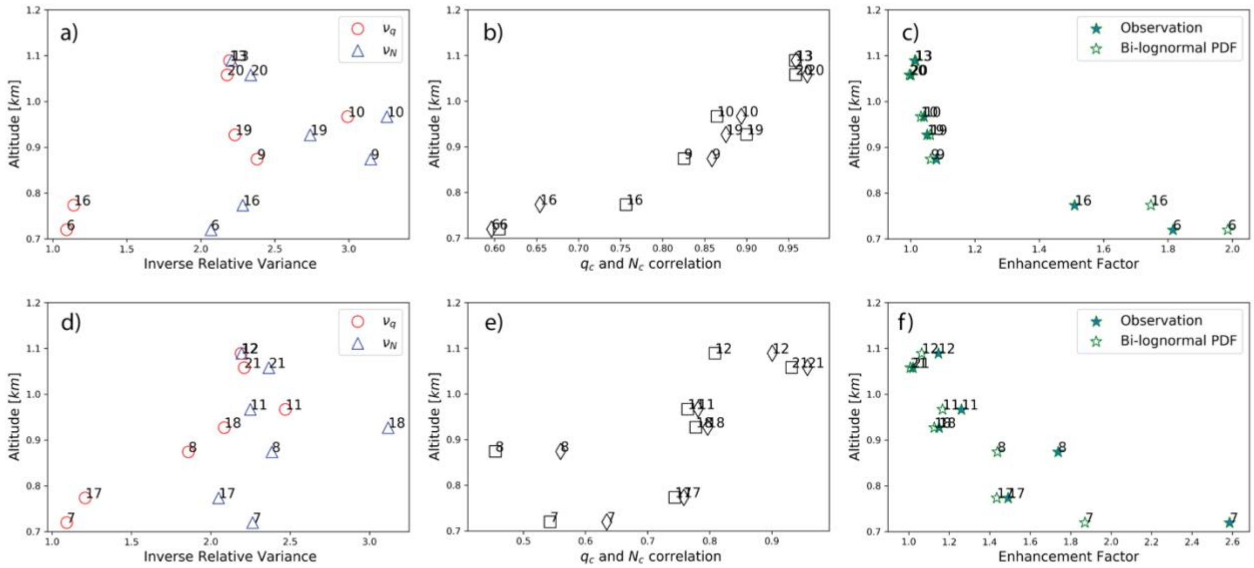

Figure 2 a) vertical variations of a) ν_q and ν_N b) q_c and N_c correlation, and c) the total EF for the new hlegs 6, 9, 10, 13, 16, 19, 20 (west side of the "v" shape flight); d) e) and f) are same as a), b) c), respectively, except or for the new hlegs 7, 8, 11, 12, 17, 18, 21 (east side of the "v" shape flight).

2) Flying in and out of cloud?
How are data for those legs where the aircraft goes in and out of cloud being handled? I assume that you didn't include a lot of zeros in either qc and/or N when calculating the means and variances, or for that matter, the directly calculating E values (e.g., equation 9)? Please describe what was done and comment on the impact this might have on the results.
Reply: As explained in the section 4.1, we use q_c>0.01 gm^{-3} as the threshold to mask cloudy against clear-sky observations. All of our analyses are based on in-cloud measurements. In other words, we exclude zero values in the computations of ν_q, ν_N, E_q, E_N or E, etc. For example, the $P(q_c)$ in Eq. (6) is the in-cloud PDF of q_c in a GCM grid. We have explicitly point this out at the beginning of Section 4.1. In addition, to avoid potential confusion, we have replaced the lower limit of the integration in Eq. 3, 9 and 10 from zero to minimal in-cloud values $q_{c,min}$ and $N_{c,min}$.

3) Inner Scale Dependence
One of the most interesting points in the paper is that the PDF of N is not lognormally distributed near the boundaries. It seems to me that this is likely due to entrainment

preferentially evaporating the smallest cloud droplets first. The time series is very spikey in nature (spikes having low N values). This suggests to me that this effect might be greatly reduced if you started from a larger inner scale (100 or 200 m) rather than the ~10 m.

I think it would be a nice addition to your analysis to first average the FCDP data to a scale of 100 to 200 m (use 1 Hz data) and examine how this impacts the skewness of the distributions and the ability of the bivariate log normal to model the variability and enhancement factors.
Reply: Following your suggestion, we did a sensitivity study using 1 Hz FCDP data (instead of the original 10 Hz) for the analysis. As an illustrative example, Figure 3 shows q_c *and* N_c *time series based on 1 Hz (upper panel) vs. 10 Hz (lower panel) FCDP observations* for the hleg #8 at cloud top.

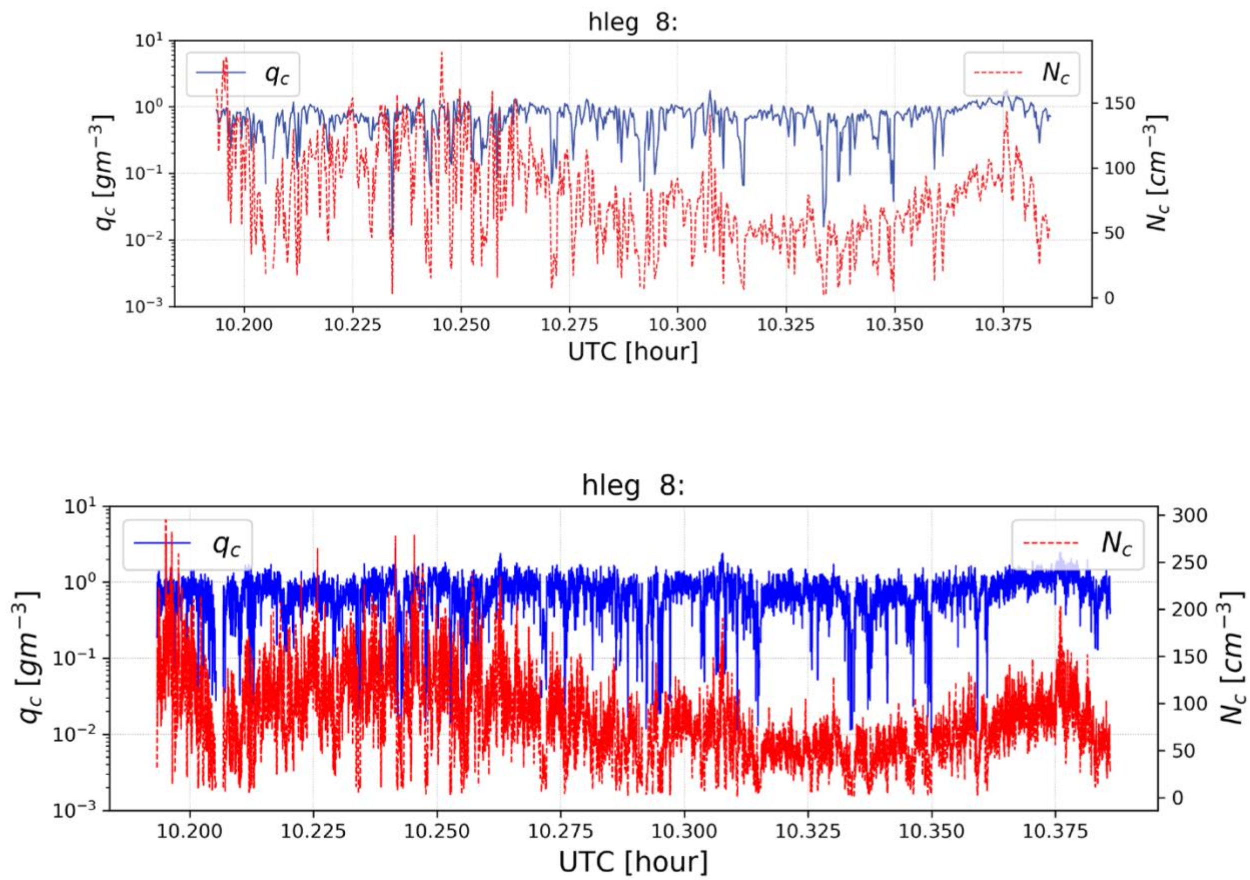

Figure 3 q_c and N_c time series based on 1 Hz (upper panel) vs. 10 Hz (lower panel) FCDP observations.

Using the 1 Hz data, we repeated all the analyses we did for the July 18, 2017 case. As shown in Figure 4 and Figure 5, the results are <u>almost identical</u> to those based on 10 Hz data. Therefore, we can conclusions hold for both 10 Hz and 1 Hz FCDP data.

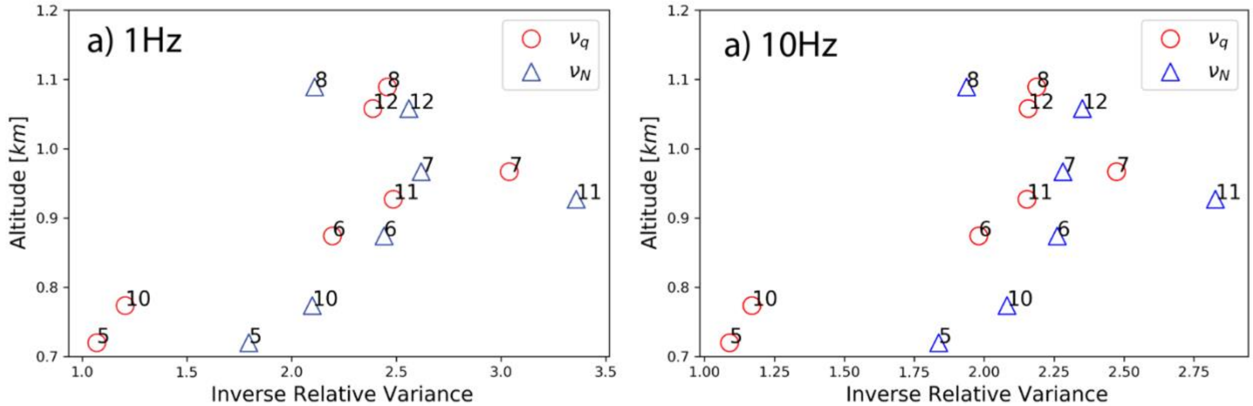

Figure 4 inverse relative variances based on a) 1Hz and b) 10 Hz

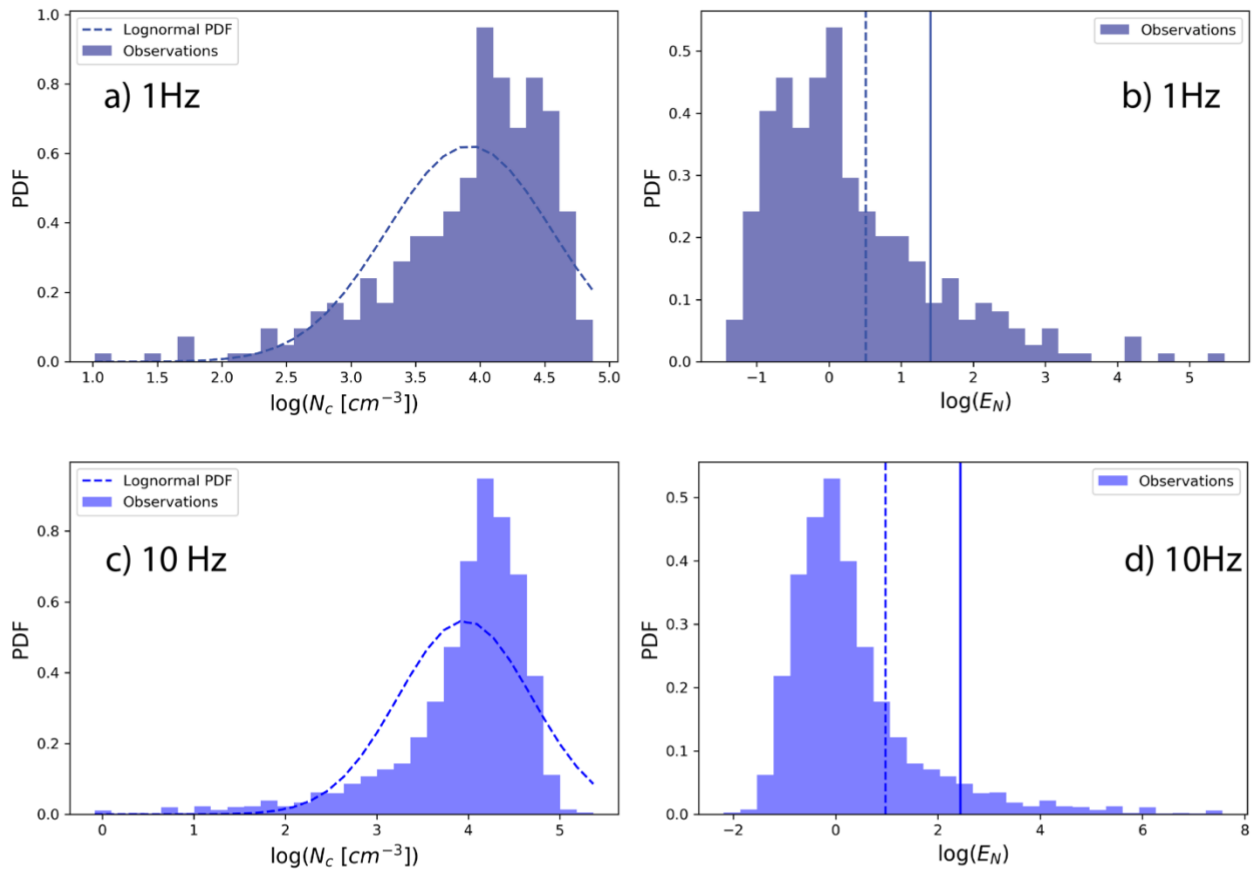

Figure 5 Analysis of the En bias for hleg 10 based on 1 Hz FCDP data (a and b) vs. 10 Hz FCDP data (c and d).

4) Dependency rather than trend

I think the discussion is overly focused on whether there is a (linear) trend in the data. Let's look at Figure 9. To me, "trend" just doesn't seem like the right word to describe the situation in Figure 9a. I suggest a more apt description might be, "Figure 9a shows significant scatter in E with altitude, with smaller values and relatively little scatter near cloud top, and decidedly larger values near cloud base."

I understand entirely that the change with height projects on to a line and the correlation is significant. There is a vertical dependence. But it is not obvious to me that there is a linear dependence or trend.

Importantly I see Figure 9c has this same pattern, with larger En values (on average) at the bottom than the top, as does 9b for Eq with the exception of the two outliers near cloud top.

Somewhat similarly in 9d, the correlations are clearly stronger and there is less variability near cloud top. But it is not clear to me there is a linear dependence. I note the bimodal points are causing much of the variability mid-cloud.

I recognize this is partly philosophical, but I think it would be better to recast some of the discussion in terms of a there being a vertical dependence rather than a "trend", though perhaps with more points (general comment 1b) a trend will become clearer.
Reply: Good point! We put all the cases together in Figure 9 in hoping to see if there is any common feature. And the only common features we could observe are the decreasing "trends" in E (Figure 9 a) and increasing "trend" in ρ (Figure 9d). But we agree with your observation that it is not very convincing to call it a trend. Indeed, the vertical variation of ν_q, ν_N, and thereby EF are unlikely going to be linear. On the other hand, because of the limited sampling rate and the big differences among the selected cases, it is hard, if not impossible, to resolve both horizonal and at the same time the detailed vertical variations of cloud properties using the in situ measurements. We revised the manuscript following your suggestions and also pointed out the caveats when interpreting the results in Figure6 and 9 (at the end of Section 4.1).

Specific Comments:
Line 213. What is the WCM-2000? Please provide a reference and perhaps add to table #1. Perhaps add a figure demonstrating agreement in qc, or at least quantify what "excellent" agreement means (e.g. Bias less than X and RMS difference less than Y% at 10 Hz?). I note here that this study is about variability and so ideally it would be good to establish variability in qc is the same from both sensor (not simply that the bias between two measurements is small).
Reply: WCM stands for "Water Content Measurement" ("2000" is just a model number). We added a reference (Alyssa and Mei) for this instrument. We compared the qc from FCDP with the WCM measurements for the selected hlegs of the July 18, 2017 case. The mean values of the two measurements are generally within 20% (this information is added to the manuscript as suggested). We have consulted the DOE measurement team (i.e., Fan Mei) and confirmed that the difference is reasonable. The two instruments seem to have a time lag, probably due to the instrument response difference. The instrument differences are beyond the scope of this study.

Table #1. Perhaps change label "Accuracy" to "Particle Size Intervals" or something similar?
Reply: Good suggestion. We changed to "size resolution".

Line 214. A value of 20 microns for ›˙IS′ § is about as small a threshold as might be chosen, and there are often a significant number of particles in the 20 to 50 micron size range for clouds that are described as "non precipitating". I expect particle concentrations in the 20 to 50 micron size range will co-vary with the concentration of particles less than 20, so I am not surprised that you see little sensitivity. Nonetheless I think it would have been better test of the sensitivity by* *choosing values for r* of 20 and 50, rather than 20 +/-5. In particular, you show later in the manuscript that N (with the cut off of r* of 20) is not well described by a lognormal near cloud-top or cloud-base. Is this result sensitive to the choice of r*?*

Reply: The r^* is the threshold used to separate the "cloud mode" and "precipitation mode", and also the autoconversion and accretion processes. As explained in the paper, we choose r^*=20 µm to follow previous studies (e.g., Wood 2005). r^*=50 µm seems too large too us. Nevertheless, we did a sensitivity study in which we set r^*=50 µm and got almost identical results (see Figure 6 below). Your expectation is correct. The results are not sensitive to the choice of r^*.

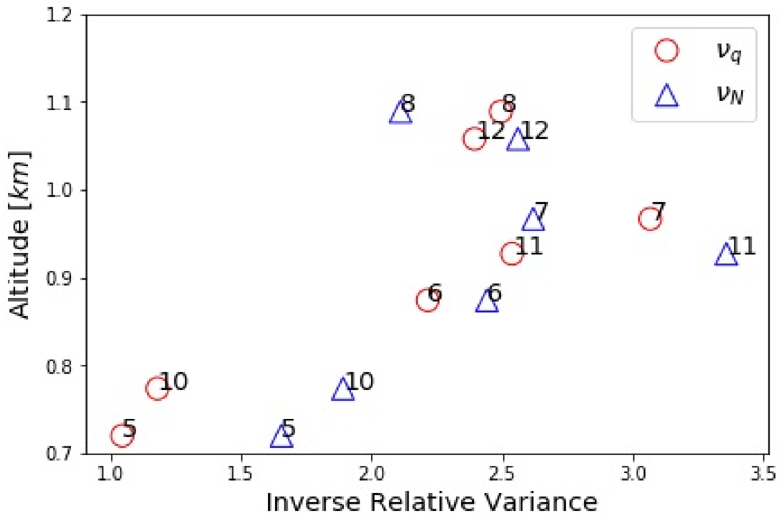

Figure 6 Inverse relative variance plot when set r^*=50 µm

Line 238. Do I understand correctly that only flight labeled as drizzling in table #2 were considered? If so, how were the groupings in Table #2 established? Does "non precipitating" in table #2 mean only particles less than 20 microns where present OR Only when drizzle is clearly falling from clouds (from radar and "pilots" on line 269)?

Perhaps it doesn't matter for this paper, but I presume models will apply autoconversion rate parameterizations all the time and I think it would be useful to examine precipitation formation for cloud without obvious precipitation / virga falling from the bottom OR otherwise clearly establish the conditions under which your results apply.

Reply: Reviewer #2 raised somewhat similar questions. At the beginning, we used the pilot summary in table #2 to narrow down our search for heavily drizzling cases. Then, we manually selected the 4 cases mainly based on radar observations (i.e., strong radar reflectivity with precipitation reaching surface). We didn't select the non-precipitating cloud cases for a couple of reasons. The first reason is that we would like to make sure that the drizzling processes, including both autoconversion and accretion have been initialized in the selected case. The second reason is more practical. It is because non-precipitating clouds are usually physically thinner than precipitating clouds, which makes it difficult for the airplane to sample different vertical locations of the clouds. As a result, there is often only one or two in-cloud hlegs for the non-precipitating clouds.

Nevertheless, per your suggestion, we selected three non-precipitating or weakly precipitating cases: 1) 2017-07-13 (non-precipitating) 2) 2018-01-26 (weakly precipitating at cloud base but no perception on the ground) 3) 2018-02-07 (very weakly precipitating at cloud base but no perception on the ground) and added them to the revised manuscript. The radar reflectivity curtain with vertical flight track for these three cases are shown in Figure 7 below. The abovementioned challenge of sampling thin non-precipitating cloud can be clearly seen in Figure 1a for the 2017-07-13 case. The selected hlegs and vlegs for these cases are summarized in Table 2. We repeated the same analyses for these new cases as for other cases, i.e., the vertical and horizontal structures of qc and Nc, as well as the EF, for these newly added cases. Overall, the results from these newly added non-precipitating cloud cases are highly similar to those based on the July 18, 2017 case as discussed in section 4. Take the 2018-02-07 case for example. Figure 8 shows the vertical variation of the inverse relative variances ν_q and ν_N. Apparently, both ν_q and ν_N demonstrate a pattern similar to that of the July-18, 2017 case (see Figure 4c of the paper), i.e., increasing first from cloud base (hleg 1 -> hleg 2) and then decrease toward cloud top (hleg 3). Therefore, these newly added cases do not affect the general conclusion although they add to the statistics.

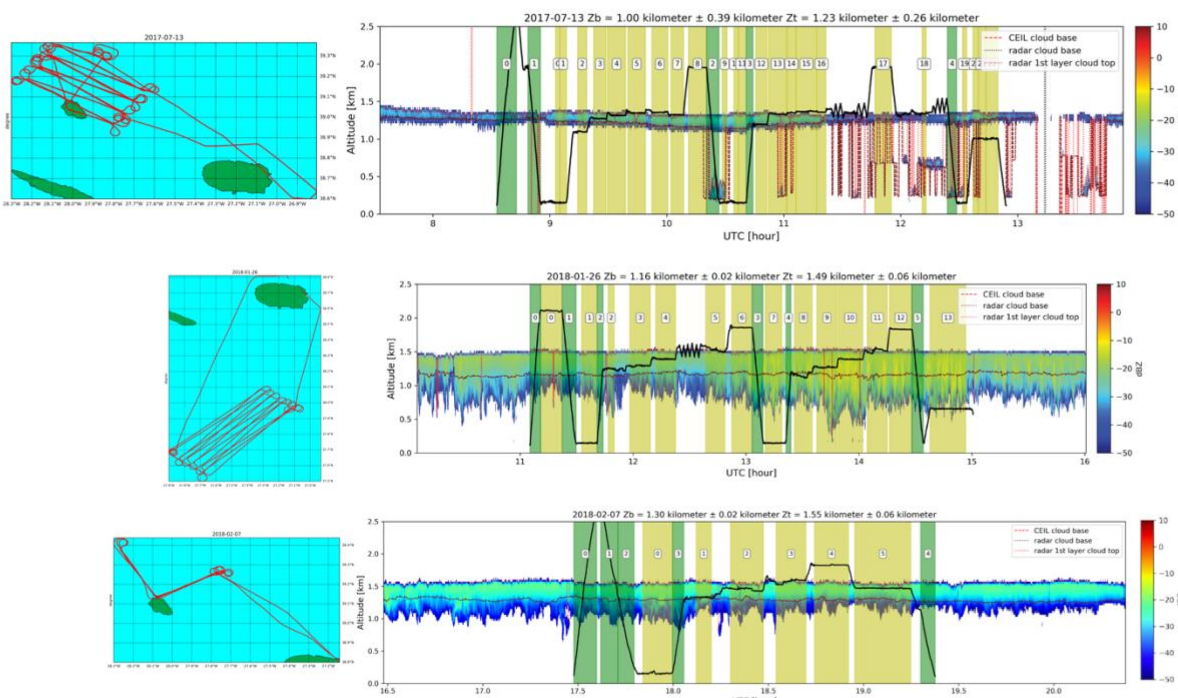

Figure 7 Three non-precipitating (or weakly precipitating) clouds added to the revised manuscript.

Table 2 A summary of selected RFs, and the selected hlegs and vlegs within each RF. (newly added non-precipitating cases are highlighted in bold font)

Research Flight	**Precipitation**	**Sampling pattern**	**Selected hlegs**	**Selected vlegs**

July 13, 2017	**Non- Precipitating**	**Straight-line**	**3, 4, 5,**	**0, 1, 3**
July 18, 2017	Precipitation reaching ground	“V” shape	5, 6, 7, 8, 10, 11, 12	0, 1, 3
Jan. 19, 2018	Precipitation reaching ground	“V” shape	6, 7, 8, 15, 16	0, 1, 3
July 20, 2017	Precipitation reaching ground	“V” shape	5, 6, 7, 8, 9, 13, 14	0, 1
Jan. 26, 2018	**Precipitation only at cloud base**	**Straight-line**	**3, 4, 5, 9, 10, 11**	**0, 1, 3**
Feb. 07, 2018	**Non- Precipitating**	**“V” shape**	**1, 2, 3, 5**	**0, 1**
Feb. 11, 2018	Precipitation reaching ground	Straight-line	4, 5, 6, 7, 12, 13	0, 1

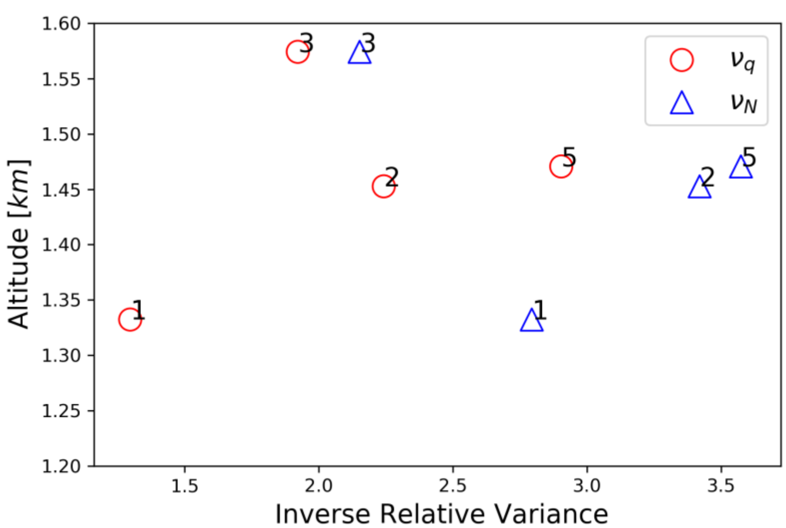

Figure 8 vertical dependence of the inverse relative variances for the Feb. 07 2018 case.

Line 269. I don’t fully understand this criteria or how it is being applied. Is this requirement applied to each horizontal leg or just for choosing cases (see comment for line 238)? If individual legs, are you focusing on the portion of the leg that occurs near the *radar? Is there a reflectivity threshold? In general, you have CDP and 2DS data and so I don't understand the choice to rely on radar or pilots.*

Reply: Since we added the non-precipitating cloud cases, this criterion *is* removed. For your information, the criterion was originally applied to research flights in table #2 rather than hlegs.

Line 274. Does "same region" here mean the hlegs must occur along roughly the same track, that is the same set of latitudes and longitudes? If yes, I don't understand the rationale for this. As long as you are sampling the same cloud field (i.e. it is reasonable to expect a representative measure of the variability) and you are using hlegs with the same total length, why do you care if the hlegs and nominally stacked?

Reply: By "*same region*" we mean that the selected hlegs should be within the same "virtual" GCM grid box. The hlegs do not have to be "*nominally stacked".* Take the July 18, 2017 case for example. All the "v"-shape legs naturally formed a "virtual" grid-box. Although the airplane repeated many "v"-shape legs, there are also long transient flights (See Figure 1a of the paper). We excluded these flights because they are outside of the "virtual" grid-box.

Line 332. Previous studies? Please provide a reference.

Reply: The following references are added: Barker 1996, Lebsock et al. 2013 and Zhang et al. 2019.

Line 443. "slight"? It is not clear to me whether this difference would or would not have a significant impact on a model simulation. Perhaps just indicate the bivariate lognormal results in an EF value that is about 0.5 larger (be quantitative rather than qualitative).

Reply: Thanks for raising this question. Actually, after examination we found a small bug in our codes (see explanation at the beginning). The Eq based on lognormal parameterization are actually very close to observation-based values, which is consistent with our previous finding in Zhang et al. 2019. The mean bias is only 0.06. this result. This information is added to the revised paper.

Line 541. It seems natural to expect qc and N will be positively correlated, since stronger updraft mean both activating more droplets and condensing more water.

Reply: Yes, this is one possible reason. Another possibility is the inhomogeneous mixing due to cloud top entrainment, which reduces the qc and Nc simultaneously. In this study, we can only speculate about the possible reasons. At the moment, we are also using LES to investigate the underlying physics.

Line 503-12. The use of a "relative hleg altitude" seem to me to be "just as prone" or even "more prone" to misinterpretation as using cloud boundaries from either radar/lidar or the vlegs. Here you are effectively demanding that the highest and lowest legs in each flight should have the same relationship regardless of how close or far they are from the actual cloud top or cloud base. I guess the question is, does it matter? Based on your 4th criteria (line 281) I would hope not.

Reply: As one can see from Figure 1 of the paper, the selected cases are quite different in terms of cloud thickness and cloud boundaries. So, to make any meaningful comparison, we have to first normalize the boundary of each case, which lead us to the use of *"relative hleg altitude"*.

We can the cloud boundaries either from in-situ or radar/lidar measurements for the normalization, although uncertainties are inevitable either way.

Due the field campaign, the pilot was actually instructed to sample the cloud base and cloud top as close as possible because these observations are very useful for studying clouds (e.g., understanding cloud top entrainment and cloud base precipitation). However, it is still extremely difficult to determine cloud boundaries from in situ measurements alone. This is why we have required that the in situ and radar/lidar measurements are largely consistent (i.e., the *4th criteria (line 281)")*

Minor Comments

Line 55-6. Replace ". . . at the spatial scales much smaller .. " with ". . . on spatial scales that are much smaller . . ."

Reply: Done

Line 56-7. Remove the phrase " . . . making the simulation of these processes in GCMs highly challenging." This is entirely redundant with the "challenging" remark on the previous sentence.

Reply: Done

Line 164. Perhaps rephrase as ". . . better understand the horizontal variations in

Reply: ok done.

Line 184. Replace "seasonable" with "seasonal"

Reply: Done.

Line 185. Perhaps add comma and remove "and" so the sentence reads " . . . the MBL (Dong et al., 2014; Rémillard 185 et al., 2012), to improving cloud parameterizations in the GCMs (Zheng et al., 2016), to validating the space-borne remote sensing products of MBL clouds (Zhang et al., 2017)."

Reply: Good suggestion. Done.

Line 189. Change "is" to "was", and perhaps simplify to read "In 2013 a permanent measurement site was established by the ARM program on Graciosa Island, and is typically referred to as the ENA site (Voyles and Mather, 2013)."

Reply: Done.

Figure 1. I presume the green blobs are islands? Perhaps explain this in the caption along with note about location of the ARM site and likewise describe what the number in white boxes and colored stripes mean in radar panels? Also I suggest higher resolution figure would be helpful here.

Reply: Yes, they are islands. Figure 1 has been revised following your suggestions and figure captions updated. Note that we have moved the figures for other selected cases to the supplementary materials so we can use high-resolution and larger figures for July 18, 2017 case.

Figure 9. The solid symbols seem a bit problematic here. I am pretty sure that in panel c, there is a red dot near cloud top that must be hidden under the other symbols – and this make me wonder if other symbols might also be hidden.

Reply: we revised the figure using open symbols to reduce overlap.

Line 537. I think you mean "Eq is larger than E", and you need to change the second occurrence of Eq in this sentence to be simply E.
Reply: yes, you are right. We corrected this.

Line 561. Perhaps change to read "... implications of subgrid variability as relates to the enhancement of autoconversion rates ..."
Reply: Thanks for the suggestion. We revised accordingly.

Line 566. Perhaps add "near cloud top" to the end of the sentence. I think it is reasonable to make the association with cloud-top entrainment but obviously, the study doesn't define entrainment zone as best I recall.
Reply: *Changed to "near cloud top"*.

Line 569. Change "we" to "our".

Reply: Done.