Response to referee #2 (in RC3)

RC: This paper describes an ice concentration retrieval based on the DARDAR CloudSat/CALIPSO combined lidar-radar retrieval. The extension of DARDAR to retrieve ice concentrations, evaluation by comparison with in situ aircraft measurements, and global distributions are discussed. Although the ice concentration retrieval seems reasonable and potentially useful, I have significant concerns with the paper in its current version. In particular, I think the validity of the retrieval in regions without both lidar extinction and radar reflectivity needs much more discussion and evaluation. Also, the use of 2D-S measurements for determining concentrations of small ice crystals is suspect at best. These issues (and others) are discussed in detail below.

AR: We are thankful to the referee for the insightful comments listed in this review. We fully agree with these two major concerns regarding the behaviour of DARDAR-LIM under single-instrument conditions and the uncertainties related to small ice concentrations by the 2D-S. The manuscript has been substantially edited, with the support of supplementary materials and an appendix, to provide further clarifications and discussions on these two points. Detailed answers to each of the referee's comments are provided below.

RC: 1. The discussion of the retrieval algorithm in section 2 implicitly assumes that both extinction from the lidar and radar reflectivity are available. The authors should make clear early in the paper that the ice concentration retrieval is dubious in cirrus that are not detected by both radar and lidar (either too optically thin for detection by the CloudSat radar or below optically thick layers where the CALIOP lidar is blocked). When only lidar backscatter or radar reflectivity are available, the ice concentration is entirely dependent on the assumed size distribution. Mean PSDs are shown in the paper, but aircraft data shows that enormous PSD temporal and spatial variability is typically prevalent in cirrus. With only lidar or radar data available, this variability cannot be captured by the retrieval.

AR: The referee is absolutely correct in that DARDAR-LIM can by impacted by the absence of either lidar or radar reflectivity (i.e., referred to as radar- and lidar-only retrievals) and that this issue should explicitly be discussed in the manuscript. Following this comment, the behaviour of the algorithm under such conditions is now discussed in Sec. 3.1 as well as in Appendix A1. These sections clarify that two aspects are important to consider:

- First, it is correct that optimal retrievals should be expected in lidar-radar conditions due to having two pieces of information available to constrain both scaling parameters of the normalized size distribution (D_m and N_0^*). When only one instrument is available, DARDAR must rely on a priori assumptions, and in particular a relation between N_0^* α_ext and the temperature. Nevertheless, DARDAR also propagates, through its optimal estimation scheme, information vertically by using lidar-radar retrievals within the same column to further constrain this relation. The quality of lidar-only and radar-only N_i estimates is therefore difficult to predict. A propagation of the operational retrieval uncertainties is now proposed in the revised manuscript (see Sec. A2 and figure S2 of the supplementary materials). Figure S2 in particular shows that errors are indeed minimum in lidar-radar conditions, about 25% against 50% in lidar- and radar-only conditions, at their respective maximum of occurence. These numbers should nevertheless be carefully accounted for because DARDAR was not designed to retrieve $\mathrm{N_i}$ and importance quantities, like the shape of the PSD through the α and β parameters, are not rigorously represented.

- However, it should also be pointed out that, despite instrumental sensitivity, it can be reasonable to expect that lidar-only $\mathrm{N_i}$ estimates can in certain conditions be more accurate than lidar-radar retrievals. Indeed, lidar-only regions are often met at cloud top where the ice clouds are optically thin. Such conditions are likely to be met by small ice crystals that have not yet aggregated, and therefore display a rather monomodal size distribution that is easier to accurately be reproduced by D05. We recall that D05 assumes a monomodal shape and our study has already shown that this is a major limitation of the current method. Under lidar-radar conditions, i.e. deeper in a thick cloud structures, the PSDs are likely to become more complex and the monomodal-shape approximation followed by D05 will not hold anymore, which leads to more uncertain retrievals. In order to clarify this point, a new figure (Fig. 3) has been added to the revised manuscript. This figure explicitly compares the in situ PSDs measured by the 2D-S (coincident with A-Train overpasses, in black) to the PSD predicted by D05 using D_m and $\mathrm{N_0^*}$ from the 2D-S (i.e. the "optimal retrievals"; in red) and the PSD actually retrieved by DARDAR-LIM (i.e. using D_m and $\mathrm{N_0^*}$ from DARDAR; in blue). It is clear that in many cases the DARDAR-LIM PSD is close to the D05 PSD, indicating enough sensitivity to properly constrain D_m and $\mathrm{N_0^*}$ in all instrumental conditions. It is also interesting to note that good agreements to the 2D-S PSDs are obtained in lidar-only conditions due to their tendencies to be monomodal with less large crystals.

Therefore, deciding on the accuracy of DARDAR-LIM $\mathrm{N_i}$ estimates in lidar-only conditions is not trivial, as it depends on the instrumental sensitivity as well as the PSD shape that are met in a given cloud parcel. The manuscript has been revised to make this more clear to the reader.

RC: 2. Page 6, lines 24-28: It would be helpful if some formal estimate of the uncertainties in Ni retrievals associated with measurement uncertainties could be provided.

AR: We agree that a formal estimates of the uncertainties on $\mathrm{N_i}$ due to instrumental error and non-retrieved parameters of the forward model in DARDAR could be useful to the reader. These were not provided in the original manuscript because, as mentioned in the previous point, DARDAR was not designed to estimate $\mathrm{N_i}$ and so some non-retrieved parameters in the retrieval algorithm that are important to $\mathrm{N_i}$ (such as the PSD small mode shape) have not clearly been considered and included for error calculation. We now propagated the Gaussian uncertainties attached to IWC and $\mathrm{N_0^*}$ in order to provide quantitative uncertainties on $\mathrm{N_i^{5\,\mu m}}$, $\mathrm{N_i^{25\,\mu m}}$ and $\mathrm{N_i^{100\,\mu m}}$, which could be considered as lower error bounds. This is now discussed in Sec. 3.1, A2 and Fig. S2 of the supplements. Complementarily, the impact of the shape parameters on $\mathrm{N_i}$ is also discussed in this section and in Fig. S3.

RC: 3. Page 6, lines 26-27: Further discussion of the the uncertainty in Ni retrieval associated with PSD shape assumption should be included. Perhaps examples could be provided as a guide.

AR: We thank the referee for this comment, we fully agree that more discussion on the impact of the PSD shape should be included. This is now discussed in Sec. 3.1, A3 and in Fig. S3 of the supplements. In this figure, several examples of PSD shapes are considered. Values of N_0^* and D_m representative of 3 temperature bins, based on the in situ campaigns used in this study, are considered as well as examples of 9 couples of α and β parameters extracted from several in situ campaigns by Delanoë et al. [2014]. It clearly appears in the top figure of S3 that these parameters indeed greatly influence the assumed PSD shape. Consequences on $N_\mathrm{i}^{5\,\mu\mathrm{m}}$, $N_\mathrm{i}^{25\,\mu\mathrm{m}}$ and $N_\mathrm{i}^{100\,\mu\mathrm{m}}$ are shown below. Considering the typical values reported by Delanoë et al. [2014], an overestimation of about 50% can reasonably be expected on $N_\mathrm{i}^{5\,\mu\mathrm{m}}$, with the exception of one sub-visible (thin cirrus) case representative of a much higher concentration in small crystals than D05. Lower uncertainties due to the PSD shape are expected on $N_\mathrm{i}^{25\,\mu\mathrm{m}}$ due to a lesser influence of the α parameter.

RC: 4. As noted in the manuscript, only 2D-S data was available from SPARTICUS. The 2D-S ice concentrations are overwhelmingly dominated by the 1st size bin (5-15 μm). Artifacts and uncertainties render the first bin or two of 2D-S measurements relatively useless. Most 2D-S data users do not use concentrations in the first two bins in their analyses because of the large uncertainties. I would recommend excluding the first two bins in the PSD comparisons shown in Figure 1 for temperature bins for which little or no ATTREX data is available. Also, I think it is inappropriate to use $N_\mathrm{i}^{5\,\mu\mathrm{m}}$ data from the SPARTICUS 2D-S-only dataset for evaluation of the satellite retrievals. The MACPEX 2D-S data should only be used for $N_\mathrm{i}^{25\,\mu\mathrm{m}}$ and $N_\mathrm{i}^{100\,\mu\mathrm{m}}$.

AR: We agree with the referee concerning the high degree of uncertainties of the ice concentrations measured in the 2D-S first bin (5-15 μm). Despite that this matter is still actively discussed, as the response to this review provided by Referee #1 (RC4) illustrates very well, it is important to be careful when dealing with this data. That is why several D_min thresholds were used in this study, including a D_min=25 μm that allowed for excluding the first two size bins of the 2D-S. The reader can then decide on the degree of trust they put on the 2D-S data and, subsequently, on the DARDAR-LIM evaluation. It should be noted that all three thresholds investigated here (D_min=5 25 and 100 μm) are part of the product that will soon be made publicly available. After careful consideration, we have decided to not completely remove $N_\mathrm{i}^{5\,\mu\mathrm{m}}$ analyses from the manuscript but the discussions were instead largely edited throughout the manuscript to remind that $N_\mathrm{i}^{25\,\mu\mathrm{m}}$ is a more trustworthy reference when it comes to 2D-S data. Further discussions on issues with 2D-S measurements in its first 2 bins have also been included in Sec. 3.2 and 3.3. Finally, all analyses (including for the case study and geographic distributions) are now extended to $N_\mathrm{i}^{25\,\mu\mathrm{m}}$. The conclusion was also edited to stress the need for more evaluation of the DARDAR-LIM $N_\mathrm{i}^{5\,\mu\mathrm{m}}$ because of these issues. We hope that this should provide enough information to the reader to make an educated choice regarding its use of the N_i dataset that will be provided co-jointly with these papers.

However, it should be noted that $N_\mathrm{i}^{5\,\mu\mathrm{m}}$ predictions by D05 agree fairly well with the FCDP and NIXE-CAPS measurements (with a possible overestimation by less than a factor of 2 but a good correlation). This gives in further confidence in that $N_\mathrm{i}^{5\,\mu\mathrm{m}}$ by DARDAR-LIM are useful and contain physical meaning, even if further investigation based on coincident flights will be required to assess the accuracy of their absolute value.

Following the referee's advice Fig. 1 was edited to exclude the two first bins of 2D-S where little SPARTICUS data is available by comparison to ATTREX (i.e., for $T_c < -60°C$). However, the MACPEX 2D-S data has not been added to this study as it would not add much more information by comparison to the dataset already used, especially considering the uncertainties on the 2D-S.

Finally, it should be mentioned that, following a comment in by Referee #1 (in RC2), the SPARTICUS dataset has been updated. The new dataset is now based on 2D-S data treated with a different processing method for $D < 365\,\mu m$ (see response to RC2 or edits in Sec. 3.2.1 of the revised manuscript). This does not change in any way the conclusions of this study but slights differences in the figures will be noted.

RC: 5. Figure 1: Indicate in figure or caption which temperature ranges correspond to ATTREX data (mostly <-70 ° C) and SPARTICUS data (warmer temperature ranges).

AR: We agree that this would be a useful information to the readers. In order to avoid including too much information in Fig. 1, an histogram showing the temperature distribution for all included campaigns is added in the supplements (Fig. S1) and is referred to in the caption of Fig. 1.

RC: 6. Figure 2: The authors should note and discuss the D05 overestimate (by factor of 2-3) for small ($D < 10\,\mu m$) particles in -80 to -70°C bin compared to ATTREX measurements.

AR: Thank you for noting this important point. It is true that D05 seems to overestimate the concentration in small ice crystals (smaller than about 15 to 25 μm) due to a too steep representation of the PSD small mode (too negative α coefficient). This is now noted and discussed in the analysis of this figure (now Fig. 1 in revised manuscript) and the now Fig. 2 of the revised manuscript, which indicates a subsequent overestimation by a factor less than 2 on $N_i^{5\,\mu m}$. This point is also now mentioned in the conclusion as it is an aspect that should be improved in future parameterizations used for N_i retrievals. It can be mentioned that the PSD parameterization planned for the next versions of DARDAR possesses a less steep representation of the concentration in small particles (i.e., a less negative α, as can be noted in Fig. S3, the new parameterization being "all (DARDAR)"). Improvements are therefore expected in future DARDAR-LIM versions, but discussing them at this stage is out of the scope of this paper.

RC: 7. Figure 3: The small sample volume of the FCDP instrument results in discretization of the ice number concentration in steps of about 12 $conc.bin^{-1}$. In other words, the FCDP instrument cannot effectively measure ice concentrations smaller than about 10-20 L^{-1} if the data is used at 1 Hz (as in this study). The CAS data has a similar sample volume issue. Since ice concentrations are often dominated by the small crystals sampled by FCDP and CAS, I would recommend not showing the in situ vs D05 comparisons for concentrations less than 10

L^{-1}. In some of the temperature bins, the data extends to ice concentrations greater than 1000 L^{-1}. Extending the upper limit on the Figure 3 axes would be helpful to show how well the retrieval compares with in situ measurements at higher ice concentrations.

AR: We completely agree with the referee that the FCDP and CAS instruments are not optimal for measuring small concentrations less than about 15 L^{-1}. However, it can be argued that in the occurence of such small N_i, the overall concentration is likely to be dominated by large particles that are not measured by these two instruments. Additionally, a minimum detection limit is used in the treatment of the CAS and FCDP data so that concentration smaller than that threshold leads to no signal. Therefore, concentrations less than 15 L^{-1} only originate from ice crystals larger than $25\,\mu$m in the 1-Hz dataset. This is not necessarily the case in our dataset as 10 1-Hz PSDs are here merged to create PSDs representative of a 10-s sampling (comparable to the CloudSat overpass, as discussed in Sec. 3.2.2). We therefore kept $1\,L^{-1}$ as the lowest value for these analyses.

Nevertheless, following this comment, the concentrations axes in Fig. 3 (now Fig. 2 of the revised manuscript) have been extended to 5000 L^{-1} to encompass all measured concentrations.

RC: 8. Figure 3: The authors should note that discrepancies up to factors of 2-3 occur but are difficult to see with the log-log axis scales.

AR: We thank the referee for pointing this out. Additional lines have been added to this figure in order to explicitly show a factor of 2 and 3 around the one-to-one line.

RC: 9. Will the Ni data be made publicly available? If so, data quality flags should be included to indicate when both radar and lidar signals are available as well as when the retrieval is questionable based on in situ comparisons?

AR: Yes, we think that it is extremely important that this dataset is made publicly available as soon as possible, hopefully together with these papers. A procedure has been initiated to distribute this N_i dataset via the ICARE data center (http://www.icare.univ-lille1.fr), next to the operational DARDAR product. Level-2 (orbital retrievals) as well as Level-3 (daily/monthly gridded means) are currently being produced and we hope to be able to announce a DOI together with the final version of the manuscript. A choice was made to wait for the end of the reviews in case of significant changes to the methodology were requested.

The L2 dataset will include $N_i^{5\,\mu m}$, $N_i^{25\,\mu m}$, $N_i^{100\,\mu m}$, as well as numerous flags that will allows to filter for instrumental conditions, cloud types and iteration numbers. It is difficult to create an additional quantitative flag that will reflect the conclusions of the in situ comparison made in this paper but the temperature (from ECMWF reanalyses) is included in the L2 dataset to provide some flexibility to the users in that direction. A filtering following what has been done for Sec. 6 of this study will be applied for the L3 climatologies.

RC: 10. Page 14, lines 1-6: I do not understand what the authors are saying here. I was under the impression that Figures 2 and 3 simply showed statistical comparisons between the in-situ-measured and retrieved PSDs and ice concentrations. The first paragraph of section 4.2 suggests the comparisons in section 4.1 were ideal cases. Perhaps this idealization should be explained and emphasized at the beginning of section 4.1.

AR: We thank the referee for noting that this point was not very clear. As indicated in the introduction and the beginning of Sec. 3.2, two main questions are investigated in this in situ evaluation. First it is determined if D05, which predicts PSDs on the basis of IWC and N_0^*, is capable of accurately predicting the concentration in small particle and therefore $\mathrm{N}_\mathrm{i}^{5\,\mu\mathrm{m}}$ and $\mathrm{N}_\mathrm{i}^{25\,\mu\mathrm{m}}$. Second, it is checked that enough sensitivity is available in lidar and radar measurements to properly constrain these two input parameters. Fig. 1-3 of the original manuscript responded to the first question by comparing in situ measurements of PSDs and N_i to equivalent predictions by D05 (obtained on the basis of IWC and N_0^* extracted from the same in situ data). In other terms, these correspond to optimal N_i estimates from DARDAR-LIM since we assume that IWC and N_0^* perfectly fit the in situ measurements, as if they were perfectly retrieved by DARDAR. This allows to disentangle the errors originating from PSD shape assumptions, which are tested here, from errors related with a lack of sensitivity in lidar-radar measurements, which are investigated later using co-incident flights. Therefore, these comparisons allow to clearly conclude on the limitations of the D05 parameterization and what needs to be improved (e.g. a better representation of the bi-modality) to obtain better N_i estimates.

This point was clarified by editing the first paragraphs of Sec. 4.1 and 4.2 in the revised manuscript.

RC: 11. Section 4.2: I am not convinced that the near-coincident in situ/satellite retrieval comparisons are useful given the enormous spatial/temporal variability in cloud properties and the corresponding need for very close time and space coincidences for meaningful comparisons. Not surprisingly, the scatter in the comparisons shown in Figure 4 is very large, spanning 1-2 orders of magnitude.

AR: We completely acknowledge that it is extremely difficult to colocate and compare aircraft and satellite measurements. Such attempts are still common to evaluated satellite products, including DARDAR [Deng et al., 2012], and always show a strong scatter in direct comparisons. However, even if 2D-S and DARDAR-LIM and not comparable one-to-one, the constraints taken here on the time and space collocation (i.e., 5 km and 30 min) should at least allows them to be statistically representative of similar cloud samples.

Fig. 4 of the original manuscript did not provide very good quantitative comparisons and so it is now moved to the supplementary materials (see Fig. S4 and S5). It can still be noted that the average agreement (around the center of the 1-σ isoline in dashed white) agrees well with the one-to-one line for $\mathrm{N}_\mathrm{i}^{100\,\mu\mathrm{m}}$ and shows an expected overestimation by a factor of about 2 to 3 for $\mathrm{N}_\mathrm{i}^{25\,\mu\mathrm{m}}$ and $\mathrm{N}_\mathrm{i}^{5\,\mu\mathrm{m}}$. This overestimation is consistent with expectations from the limited representation of the PSD shape by D05, as can be observed by comparing Fig. S4 and S5.

Instead, comparisons of the PSDs measured by the 2D-S, predicted by D05 based on 2D-S data, and retrieved by DARDAR-LIM are shown in the new Fig. 3 of the revised manuscript. This figure shows that DARDAR-LIM PSDs are very consistent with the D05 predictions, meaning that the cloud volumes sampled by the 2D-S and CloudSat/CALIPSO are statistically comparable most each temperature bins and instrumental conditions. The comparisons in terms of histograms, shown in Fig. 4 of the revised manuscript, now also include mean $\mathrm{N_i}$ values to allow for a more quantitative statistical comparisons.

Section 4.2 has therefore been substantially edited to include and adapt to these new analyses, which should provide more convincing evidence of the statistical comparability of 2D-S and coincident DARDAR-LIM products.

RC: 12. Page 15, line 9-10: In contrast to the statement here, the DARDAR-LIM retrieval overestimates $\mathrm{N_{D>5\mu m}}$ and $\mathrm{N_{D>25\mu m}}$ compared to SPARTICUS data even in the -60 to -50° C temperature bins.

AR: We thank the referee for pointing this out. The corresponding sentence has now been removed from the revised manuscript because this section has been substantially edited, following the response to the previous point. As a response to this comment, lines have been added in Fig. 4 of the original manuscript (now Fig. S4) in order to explicitly show a factor of 3 around the one-to-one line and identify more clearly these overestimations.

RC: 13. Figure 5: The comparisons shown here are very difficult to see, particularly those for lidar-only and radar-only retrievals. The relative agreement between lidar-radar, radar-only, and lidar-only retrievals should be shown in a separate figure, particularly since the lidar-only and radar-only retrievals are suspect. Also, as discussed above, the SPARTICUS 2D-S-only ice concentrations for D>5 μm are dominated by the first size bin, with enormous associated uncertainties. The comparisons with SPARTICUS 2D-S-only ice concentrations including the first bin are of little value, possibly misleading, and should be removed.

AR: We acknowledge that Fig. 5 of the original manuscript (now also Fig. 5) did not provide clear quantitative messages regarding the differences between 2D-S, D05 and DARDAR-LIM. These were even more difficult to observe for lidar-only and radar-only conditions as often less retrievals are available. We have responded to this issue by included the values of geometric means of $\mathrm{N_i^{5\,\mu m}}$, $\mathrm{N_i^{25\,\mu m}}$ and $\mathrm{N_i^{100\,\mu m}}$ for 2D-S, D05 and DARDAR-LIM. The overall mean values are shown, as well as the values corresponding to $\mathrm{N_i}$ estimates obtained in lidar-, radar-only and lidar-radar conditions. Also, as advised by the referee, individual histograms for each of these conditions are shown in Fig. S6 to provide a more clear idea concerning the influence of different instrumental conditions on the retrievals.

Regarding the evaluation of $\mathrm{N_i^{5\,\mu m}}$, we have chosen to keep the corresponding plots in this analyses for the reasons discussed in the response to point 4. However, we fully agree with the referee that great care should be taken when presenting results using concentration from the first size bin of the 2D-S. Sec. 4.2 has therefore been edited so that its analyses are more centered on $\mathrm{N_i^{25\,\mu m}}$ and to contain an explicit warning to the reader that $\mathrm{N_i^{25\,\mu m}}$ constitutes the better reference concentration for the 2D-S. The revised conclusion also repeats this message as a reminded that more evaluation of $\mathrm{N_i^{5\,\mu m}}$ remains necessary.

RC: 14. Page 19, lines 8-10: the lack of clear transitions in retrieved properties between the lidar-only, lidar-radar, and radar-only regions does not necessarily mean the lidar- only and radar-only Ni retrievals are credible.

AR: We agree with this comment, a lack of transition between lidar/radar-only and lidar-radar regions does not necessarily prove the quality of $\mathrm{N_i}$ estimates obtained in single-instrument conditions. This sentence was more meant as an observation rather than a definite proof, and has therefore been toned down in the revised manuscript. This sentence is also more justified now that the quality of lidar- and radar-only $\mathrm{N_i}$ estimates is further discussed in the revised Sec. 4.2. Nevertheless, this lack of transition is still worth commenting on as it represents a very impressive feature of the DARDAR algorithm, which allows for a real multispectral consistency between it's lidar and radar retrievals. It also shows that some information is used to constrain the $\mathrm{N_i}$ estimates as there do not seem to jump back to some a priori value.

RC: 15. Figure 8: Scatter plots of $\mathrm{N_{D>5\mu m}}$ and $\mathrm{N_{D>100\mu m}}$ versus $\mathrm{N_{2D-S}}$ would provide much clearer comparisons between the retrievals and measurements. Further, the points could be color coded to indicate whether they are in the lidar-only, lidar-radar, and radar-only regions. In the discussion of Figure 8, the authors claim good agreement between the in situ and retrieved ice concentrations, and they dismiss glaring discrepancies as being caused by the imperfect time coincidence. This argument seems unjustified. The flight track segment has been chosen for good time/space coincidence.

AR: We thank the referee for this comment, it has allowed us to realize that the in situ analyses included in the case study was perhaps not clearly explained. Because of this, and following a comment in RC1 this figure has been moved to supplementary materials (see Fig. S7) and is now only briefly mentioned. This is also justified since this figure mainly supported the previous conclusions, with redundant results. We however think that this figure can in this format still provide good insights on the quality of DARDAR-LIM as it shows that the satellite estimates are to some extent well capable of reproducing the spatial variability in $\mathrm{N_i^{5\,\mu m}}$, $\mathrm{N_i^{25\,\mu m}}$ and $\mathrm{N_i^{100\,\mu m}}$ measured by the 2D-S. This is why scatterplots are not proposed for this figure, especially since scatterplots have already been widely examined before. The added-value of figure is to show a consistency in the spatial distribution along the aircraft fight.

We nevertheless agree with the referee's comments and have added information on the instrumental conditions met in DARDAR-LIM in the new figure (see color background in Fig. S7). The $\mathrm{N_i^{25\,\mu m}}$ is also included in this study, as it now represents the new reference for small ice concentration from the 2D-S. Finally, the part of the flight leg that was further away from the satellite overpass (more than 10 km) has been removed and the coordinates have been changed so that the reader can easy grasp the temporal and spacial distance between the aircraft track and the satellite overpass.

We agree that discrepancies appear, in particular in the $\mathrm{N_i^{25\,\mu m}}$ comparisons and during the descending leg, which we attributed to the distance (about 10 km) from the track. The hypothesis of different cloud sampling between the satellite and the aircraft appears reasonable especially since different $\mathrm{N_i^{100\,\mu m}}$ values are noted during this descending leg. We recall that no serious issues are expected in the DARDAR-LIM $\mathrm{N_i^{100\,\mu m}}$ estimates based on the previous in situ evaluations and note that similar increases of $\mathrm{N_i^{100\,\mu m}}$ can be observed in Fig. 6(h) (of the revised manuscript) right next to the descending part of the track. It is therefore reasonable to attribute issues in $\mathrm{N_i^{100\,\mu m}}$ to the sampling different cloud volumes, which means that differences in $\mathrm{N_i^{5\,\mu m}}$ and $\mathrm{N_i^{25\,\mu m}}$ could as well be expected in this area.

RC: 16. Section 5.3: The authors describe a cloud formation scenario with air parcels ascending across the -40°C isoline, which suggests that freezing of liquid drops could be the main ice formation mechanism. Yet they attribute the differences between the high and low ice concentration regions to differences in vertical wind speed and cite the strong sensitivity of Ni to w (citing Krämer et al. 2016; papers showing this sensitivity decades earlier should be cited). However, the strong sensitivity to w occurs primarily when aqueous aerosols freeze, not so much when liquid droplets freeze. Either the description is not clear, or the physical argument made does not make sense.

AR: We are thankful to the reviewer for pointing this out. It appears that the explanation proposed in the original manuscript was misleading. We indeed meant that the relation between $\mathrm{N_i}$ and w, which we show via the analysis of back-trajectories, is the result of homogeneous freezing events of aqueous aerosols. These events are however likely to occur on top of existing ice crystals formed from liquid droplets, as it is clear that supercooled layers appear close to the region of high $\mathrm{N_i^{5\,\mu m}}$ and $\mathrm{N_i^{25\,\mu m}}$ (seen in the β_{ext} profile). We have edited this paragraph and added references and comparisons to studies that analysed these processes [e.g. Kärcher and Ström, 2003, Kärcher et al., 2006].

RC: 17. Figures 9 and 10: The discrepancy between $\mathrm{N_{D>5\mu m}}$ and ATTREX FCDP ice concentrations noted above is apparent in the coldest temperature bins and the TTL. Typical values of ND>5 μm are 200-300 $\mathrm{L^{-1}}$, whereas ATTREX in situ measurements indicate ice concentrations of about $100\,\mathrm{L^{-1}}$ (Jensen et al., 2016). It is also interesting that the ice concentrations are higher over continental and convective regions even in the coldest temperature bins (near the tropical tropopause) where the vast majority of clouds form in situ. Additionally, it might be worth noting that the statistics must be poor in the coldest temperature bin poleward of about 30° latitude since such cold temperatures rarely occur there.

AR: We are grateful to the referee for pointing this out and relating these analyses to the observed $\mathrm{N_i^{5\,\mu m}}$ discrepancies between D05 and the FCDP at very low temperatures. The section has been edited to note these, as well as the disagreements between concentrations observed in these figures and the findings of Jensen et al. [2013, 2016] for TTL cirrus. These could indeed be caused by a misrepresentation of the small ice concentration in D05, which seems to overestimate the steepness of the concentration towards small particles at low temperatures and therefore overestimates of $\mathrm{N_i^{5\,\mu m}}$ (by a factor less than 2). Interestingly, the spatial distributions of $\mathrm{N_i^{25\,\mu m}}$ (less impacted by shape assumptions), now added to Fig. 7 of the revised manuscript, indicates concentrations of about $100\,\mathrm{L^{-1}}$ in TTL regions. It can also be noted that $\mathrm{N_i^{5\,\mu m}}$ in the same regions and during winter months (see Fig. 8 of revised manuscript), i.e. when in situ clouds should be even more dominant, are about 50% lower. It therefore appears difficult to strongly conclude on disagreements by comparison to in situ observations without an exact knowledge of what cloud type is present in each lat-lon-$\mathrm{T_c}$ bin, and further analyses will be needed to fully assess this disagreement with Jensen et al. [2013, 2016]. This discussion has now been summarized in Sec. 6.1 of the revised manuscript.

We also agree that additional information on the statistical significance of the results provided here would be useful. Fig. S8 and S9 of the supplementary material now indicate the pixel counts corresponding to the spatial and zonal distributions analysed in Sec. 6.

RC: 18. Page 22, line 7: Simply stating that the spatial distributions agree with the global model predictions is no doubt too strong. A quick examination shows there are some regions of agreement and some glaring discrepancies. I would either omit this statement or qualify it. Perhaps the comparison really shouldn't be discussed without providing much more detail.

AR: We thank the referee for this comment and fully agree with it. Further comparisons to modeling is beyond the scope of this study and this sentence has been deleted from the manuscript.

RC: 19. Section 6.2: Most of the speculations about the physical causes of the zonal-height distributions in this section are either not justified or would require much more detail to adequately discuss. It does not look to me like there is a particularly sharp transition at -40°C, nor would one be expected given the importance of sedimentation in cirrus. The retrieval probably doesn't work well in the antarctic wintertime stratosphere since PSCs are typically mixtures of ice crystals, NAT particles, and ternary aerosols.

AR: We agree with this point, Sec. 6.2 contained some analyses that remained too hypothetical, such as the attribution of some $\mathrm{N_i}$ patterns to PSCs. After further investigation, it appears that DARDAR retrievals in these regions are highly uncertain due to potential failures in the cloud mask and wrong categorizations of cloud pixels. This does not mean that DARDAR retrievals are always wrong in these regions but further investigation based on the new version of the DARDAR mask should be performed and are out of the scope of this study. It is now clearly stated in the manuscript that this feature in $\mathrm{N_i}$ should not be trusted. Regarding the vertical transition at -40°C, it could still be argued that $\mathrm{N_i^{5\,\mu m}}$ and $\mathrm{N_i^{25\,\mu m}}$ values quickly change around this temperature, especially in the tropics. $\mathrm{N_i^{25\,\mu m}}$ for instance increases from about $50\,\mathrm{L^{-1}}$ to above $100\,\mathrm{L^{-1}}$. We nevertheless agree that this transition cannot be considered sharp and have toned down this analysis.

Following this comment, Sec. 6.2 has been edited to provide further discussion and remove analyses that seemed too far-fetched. The analysis of seasonal variations of $\mathrm{N_i}$ are now also supported by Fig. 8 of the revised manscript, which shows seasonal variability in spatial distributions. It should also be noted that this section also comes as a natural transition between part 1 and part 2, which further investigates some of the observations made in Sec. 6.2. This is now also reminded.

References:

J. Delanoë, A. J. Heymsfield, A. Protat, A. Bansemer, and R. J. Hogan. Normalized particle size distribution for remote sensing application. *J. Geophys. Res*, 119(7):4204–4227, 2014. doi: 10.1002/2013JD020700. URL `http://dx.doi.org/10.1002/2013JD020700`.

M. Deng, G. G. Mace, Z. Wang, and R. P. Lawson. Evaluation of several a-train ice cloud retrieval products with in situ measurements collected during the sparticus campaign. *J. Appl. Meteor. and Clim.*, 52(4):1014–1030, 2012. doi: 10.1175/JAMC-D-12-054.1.

E. J. Jensen, G. Diskin, R. P. Lawson, S. Lance, T. P. Bui, D. Hlavka, M. McGill, L. Pfister, O. B. Toon, and R. Gao. Ice nucleation and dehydration in the tropical tropopause layer. *Proceedings of the National Academy of Sciences*, 110(6):2041–2046, 2013. doi: 10.1073/pnas.1217104110.

E. J. Jensen, R. Ueyama, L. Pfister, T. V. Bui, R. P. Lawson, S. Woods, T. Thornberry, A. W. Rollins, G. S. Diskin, J. P. DiGangi, and M. A. Avery. On the susceptibility of cold tropical cirrus to ice nuclei abundance. *J. Atmos. Sci.*, 73(6):2445–2464, 2016. doi: 10.1175/JAS-D-15-0274.1.

B. Kärcher and J. Ström. The roles of dynamical variability and aerosols in cirrus cloud formation. *Atmos. Chem. Phys.*, 3(3):823–838, 2003. doi: 10.5194/acp-3-823-2003.

B. Kärcher, J. Hendricks, and U. Lohmann. Physically based parameterization of cirrus cloud formation for use in global atmospheric models. *J. Geophys. Res. Atmos.*, 111(D1), 2006. doi: 10.1029/2005JD006219.