Review of 'CALIPSO (IIR-CALIOP) Retrievals of Cirrus Cloud Ice Particle Concentrations' by Mitchell et al.

This manuscript describes and evaluates a new method dedicated to retrieve the ice particle number concentration N from satellite measurements. The method presented here infers N by using the sensitivity of a couple of thermal infrared channels to the concentration in small particles. It is applicable to cirrus with an optical depth between 0.3 and 3 and a base temperature T < 235°C. The method notably depends on several relationships between the cloud effective absorption coefficient ratio β_{eff} and diverse ice cloud optical and microphysical properties, that the authors deduce from in situ measurements. Its results are first evaluated through comparisons to in situ measurements of N and two years of global retrievals are then analyzed. Uncertainties on the retrievals are also discussed. The authors show that, despite large uncertainties on the absolute N values, the spatial variability observed in global distributions are consistent with expectations based on modeling, in situ and/or remote sensing studies.

I find this retrieval method, which provides a new yet simple way to interpret and extract information from thermal infrared measurements, very innovative and interesting. There are of course shortcomings, such as the very low ice cloud sampling after filtering or the strong dependence on relations obtained from limited in-situ data, but the authors properly discuss most of these in detail and account for them in their discussions/conclusions. This study can be seen as preliminary results of a method that could easily be improved in the future (by including more in situ data in its framework). This type of work is necessary considering that very few methods today allow for retrieving N from satellite despite the importance of this parameter to study cloud processes, understand aerosol-cloud interactions or evaluate models. The manuscript is well written but I find its length to be a serious issue. This study indeed covers a thorough technical description of the method, very detailed evaluation against in situ measurements, an analysis of global retrievals, comparisons to existing methods and even discussions on possible perspectives/applications. 51 pages from the abstract to the appendix, and 17 figures that are often multi-panel, is (even in an ACDP format) too much for this type of paper. I strongly encourage the authors to shorten the manuscript to not overwhelm the readers, who may then miss some important points. Overall, I therefore recommend for publication of these results, providing revisions following the comments listed below.

General comments:

1. A major concern with the current manuscript is its length. The authors cover a very wide range of information and results that are all interesting but their amount is, in my opinion, counterproductive. It should be kept in mind that this study will be of high interest to diverse communities, as it contains are interesting aspects for remote sensing, in situ measurements, modeling or ice microphysics. Based on the current version I suspect that readers will only focus on their section of interest and miss important results. One idea could be to shorten section 2, since a big part of it has already been described in Mitchell et al. [2016]. The results related to the $N(D)_1 = 0$ analyses are very important and should be mentioned in the paper but several of them could be moved to supplementary materials in order to lighten the discussions and the figures. Other analyses that are not focused on N (the novelty of the paper), such as section 5.4, could also be moved to supplementary materials. This would allow to remove some technical discussions and focus the paper on result analyses, which are also better suited for a publication in ACP.

2. The authors acknowledge in the manuscript that this new retrieval method is not expected to provide accurate absolute values of N. Only the spatial variability of this parameter is well represented. The accuracy of the absolute N is discussed throughout sections 3 and 4 but would it be possible to summarize a final estimate of the uncertainties on N in the conclusion? By this I mean not only the $\Delta N/N$ but also combining what has been learned from the in situ evaluation. For instance, would it be a factor of 2, 5, one order of magnitude? Also, in what conditions can optimal retrievals be expected? This type of information would be very useful to future users of this dataset, especially since this paper will serve as a reference..

3. A clear limitation of the method is its necessity to filter out a lot of data based on different cloud conditions. This, as shown in table 4, leads to samples that are only representative of up to a 2% frequency of occurrence. Is there a particular reason for not using more data? The method seems to be rather straightforward to apply on CALIPSO satellite products, and some of the co-authors should be familiar with such operational treatments. It therefore shouldn't take long to process 10 years of data and greatly improve the statistical significance of the results shown in this study (especially in sections 5 and 6).

4. Concerning the results shown in section 4 (comparisons to in situ measurements), more direct comparisons between SPARTICUS measurements and CALIPSO could be obtained by looking at the numerous co-incident overpasses between the satellite and the aircraft. This would avoid issues related to non-representative events in the in situ data by comparison to a long-term regional dataset, as noted in page 22. Have you tried looking at these exact satellite overpasses only?

5. Perhaps I have missed this information but I am under the impression that the D_{eff} retrieved by IIR is not used in this method. How different would be the results if this operational retrieval was used instead of the in situ relationship? Is there a reason for not doing this?

Specific comments:

1. Section 2.2: Computing IIR radiances, or converting between optical properties, requires some assumptions on the shape of ice crystals and of the ice particle size distribution. Dubuisson et al. [2008] for instance showed that the IIR brightness temperatures are sensitive to both parameters. Garnier et al. [2012] showed that converting the IIR effective optical depths into absorption optical depths can also depends on ice particle shape assumptions. Do you have an idea if these assumptions have an impact on your N retrievals? Are they accounted for in the uncertainties described in the appendix? Also, are the ice crystal shapes used in the MADA method to compute the in situ relationships consistent with the shapes assumed in the treatment of CALIPSO measurements, and if not is there an impact on the N retrieval and their uncertainties?

2. p. 6 l. 22: By single-layered cloud do you refer to the absence of other ice clouds or also to the absence of a liquid or mixed-phase layer underneath? IIR seems to be able to deal quite well with multi (ice+liquid) layer conditions by adjusting the background radiance. For instance Fig. 10 in Sourdeval et al. [2016] shows that IIR is robust to multi-layer conditions. It would be worth checking if removing this filter makes a big difference in your dataset. Otherwise, not excluding multi-layer scenes would clearly help to increase the global statistical representativity of N retrievals.

3. Sec. 2.4: Can error bars be added to the points in Fig. 2-4? It seems like these instrumental and computational errors could easily encompass the changes due to the choice of including the first size bin or not.

4. Fig. 9: Could you comment on the very wide spread (75-25% percentiles) in the Krämer et al N/IWC vs Tc figure by comparison to what is noted for the CALIPSO data? Is there no spread plotted for the Krämer dataset on the right plots or is it too small to be seen? Also please indicate in the caption if the CALIPSO retrievals correspond to the entire 2008 and 2013 periods.

5. Fig. 10e: As a remark, the slightly negative relation between N and T (N decreasing towards low temperature) indicated in the Krämer et al. [2009] study is not found anymore in revised version of the dataset (not yet published but seen in recent conference presentations by M. Krämer et al). This should strengthen the statement p. 31 l. 4 that differences are likely to be due to different cloud sampling.

6. Sec. 5.4: Is there an explanation to the fact that Re has a very different dependence on Tc and Tc - Ttop by comparison to what was previously shown for N?

7. Section 6.4: p. 45 l. 11-13: Another possible explanation to the differences in absolute numbers could be that DARDAR-LIM ignores the concentrations of ice particles smaller than 5 μm. p. 45 l. 14-16: As mentioned before, the slight decrease of N towards low T as noted by Krämer et al. [2009] is not found in the most recent version of their dataset, which is not anymore inconsistent with the relation shown in Gryspeerdt et al. [2018]. An increase of N towards low T is also consistent with what is shown in Figures 10ab of this manuscript. Overall, comparisons between the N retrievals presented in this study and DARDAR-LIM are very difficult as both methods are based on different approaches and difference instruments. DARDAR-LIM also retrieves vertical profiles of N whereas IIR retrievals correspond to weighted N average values from cloud top. Nevertheless, it is quite remarkable that despite all these differences the two dataset show such similar results. This clearly strengthens the confidence in both satellite products.

8. Fig. 17: It would be interesting to see the spatial distribution of the frequency of occurrence of retrievals and of the distance from cloud top (in terms of temperature) corresponding to this figure.

9. p. 49 l. 9-10: The “four formulations” have not yet been mentioned in the conclusion. It would be useful to briefly describe again in what they differ.

<u>Technical corrections:</u>

1. p. 2 l. 13 and 16: it would be better to explicitly refer to “ice clouds” instead of “clouds”

2. p. 11 l. 17: “limit is 1.031”

3. p. 43 l. 15: “m.s^{-1} in”

References

P. Dubuisson, V. Giraud, J. Pelon, B. Cadet, and P. Yang. Sensitivity of thermal infrared radiation at the top of the atmosphere and the surface to ice cloud microphysics. *J. Appl. Meteor. and Clim.*, 47 (10):2545–2560, 2008. doi: 10.1175/2008JAMC1805.1.

A. Garnier, J. Pelon, P. Dubuisson, M. Faivre, O. Chomette, N. Pascal, and D. P. Kratz. Retrieval of cloud properties using CALIPSO imaging infrared radiometer. Part I: Effective emissivity and optical depth. *J. Appl. Meteor. and Clim.*, 51(7):1407–1425, 2012. doi: 10.1175/JAMC-D-11-0220.1.

E. Gryspeerdt, O. Sourdeval, J. Quaas, J. Delanoë, and P. Kühne. Ice crystal number concentration estimates from lidar-radar satellite retrievals. part 2: Controls on the ice crystal number concentration. *Atmos. Chem. Phys. Discuss.*, 2018:1–25, 2018. doi: 10.5194/acp-2018-21.

M. Krämer, C. Schiller, A. Afchine, R. Bauer, I. Gensch, A. Mangold, S. Schlicht, N. Spelten, N. Sitnikov, S. Borrmann, M. de Reus, and P. Spichtinger. Ice supersaturations and cirrus cloud crystal numbers. *Atmos. Chem. Phys.*, 9(11):3505–3522, 2009. doi: 10.5194/acp-9-3505-2009.

D. L. Mitchell, A. Garnier, M. Avery, and E. Erfani. Calipso observations of the dependence of homo- and heterogeneous ice nucleation in cirrus clouds on latitude, season and surface condition. *Atmos. Chem. Phys. Discuss.*, 2016:1–60, 2016. doi: 10.5194/acp-2016-1062.

O. Sourdeval, L. C. Labonnote, A. J. Baran, J. Mülmenstädt, and G. Brogniez. A methodology for simultaneous retrieval of ice and liquid water cloud properties. Part 2: Near-global retrievals and evaluation against A-Train products. *Quart. J. Roy. Meteor. Soc.*, 142(701):3063–3081, 2016. doi: 10.1002/qj.2889.