RESPONSES TO THE REFEREE #2 COMMENTS

We thank the referee for his/her comments of this paper, and for constructive comments that have significantly improved this paper. We understand that this is a serious undertaking that requires considerable time and effort, and your efforts are appreciated!

In the text below, black font is used for the referee comments (RC) and blue font is used for author comments (AC) and new text added to the paper.

General comments:

RC The manuscript is quite long. If this is an issue, one could consider splitting up the retrieval/radiative aspects and the interpretation of the results.

AC We appreciate the referee's recognition of the substantial research effort that produced this manuscript. In regards to the manuscript length, we created another file titled Supplementary Materials that contains 12 figures and some of the text from the original manuscript, which shortened the paper by 8 pages.

Specific comments:

RC S1. Section 2.2.0. How do you handle the lapse through the cloud. Cirrus can be extensive in their vertical dimension. Are you errors larger for geometrically thick cloud?

AC S1. In Eq. (3), the lapse through the cloud is handled in the determination of the cloud blackbody radiance, R_{BB}. The associated blackbody temperature, T_{BB}, is determined using the approach detailed in Sect. 3 of Garnier et al. (2015) and summarized in Sect. 2.2.3 of this paper. First, we compute the IIR weighting profile in the cloud layer using the CALIOP extinction profile. The IIR weighting profile includes an attenuation term corresponding to the overlying infrared absorption optical depth (see Eq. (11) and Eq. (12) of Garnier et al. (2015)). Then, R_{BB} is the weighted averaged blackbody radiance computed using this IIR weighting profile and the GMAO temperature profile in the layer. We actually forgot to precise that the GMAO temperature profile is used, and the 3rd sentence in Sect. 2.2.3 now reads (changes in bold):

"The CALIOP lidar 532 nm extinction profile in the cloud is used to determine an IIR weighting profile that is used, **together with the GMAO GEOS5 temperature profile**, to compute R_{BB} as the weighted averaged blackbody radiance."

RC S2. Section 2.2.0: You mentioned you used the 0.55 micron extinction derived mean cloud temperature. I would expect that the cloud weighting functions are a bit different for the 10.6 and 12 micron observations and these may also differ from each other. Does this matter? Or this effect absorbed in the Blackbody radiance calculation in 2.2.3?

AC S2. For this study, T_{BB} is taken identical at 10.6 and 12 µm, following the same approach as in the Version 3 algorithm, but is improved compared to Version 3 by using the corrections detailed in Garnier et al. (2015) and summarized in Sect 2.2.3. Thus, T_{BB} is computed at 12 µm by taking the ratio, r, between visible extinction optical depth and absorption optical depth at 12 µm, $\tau_{abs}(12.05\mu m)$, equal to 2.

We agree that T_{BB} could have been estimated separately at 12 µm (noted $T_{BB,12}$) and at 10.6 µm (noted $T_{BB,10}$). Furthermore, taking the same ratio, r, for both channels means taking $\tau_{abs}(12.05\mu m) = \tau_{abs}(10.6\mu m)$ or $\beta_{eff}=1$, which is not consistent with our findings. In order to assess the error resulting from our simplified approach, we re-computed $T_{BB,12}(r_{12})$ and $T_{BB,10}(r_{10})$, with r_{12} and r_{10} not taken equal to 2, but computed respectively using $\tau_{abs}(12.05\mu m)$ and $\tau_{abs}(10.6\mu m)$ initially reported in the operational Version 3 products. The analysis was conducted over oceans in JJA 2013 between 82°S and 82°N.

- a) We find that the difference $T_{BB,12}(r_{12})-T_{BB,12}(r_{12}=2)$ (see Table 1 below) is smaller than 0.09 K on average with a mean absolute deviation smaller than 0.13 K. The resulting error is negligible compared to the assumed uncertainty of ± 2K in $T_{BB,12}(r_{12}=2)$.
- b) We find that the difference $T_{BB,10}(r_{10}) T_{BB,12}(r_{12})$ (see Table 2 below) is smaller than 0.12 K on average with a mean absolute deviation smaller than 0.07 K.
- c) Using $T_{BB,10}(r_{10})$ and $T_{BB,12}(r_{12})$ instead of the same temperature $T_{BB}=T_{BB,12}(r_{12}=2)$ as in this study reduces β_{eff} by less than 0.001 on average, with a mean absolute deviation smaller than 0.0007, which is negligible (see Table 3).

We added the following sentence at the end of the 1st paragraph in Sect. 2.2.3:

"Computing T_{BB} at 10.6 µm and at 12.05 µm yields temperatures that differ by less than 0.15 K on average, which has a negligible impact on β_{eff} for our cloud selection."

Temperature	Samples	Min	Max	Median	Mean	Standard	Mean	
T _c (K)	Count					deviation	absolute	
							deviation	
188	73	0.012	0.082	0.063	0.061	0.017	0.013	
192	348	-0.038	0.501	0.056	0.071	0.066	0.044	
196	2149	-0.685	0.486	0.076	0.087	0.085	0.056	
200	6915	-0.719	1.879	0.083	0.084	0.129	0.081	
204	16783	-1.098	3.048	0.063	0.061	0.160	0.100	
208	28620	-1.246	2.181	0.048	0.043	0.182	0.118	

Table 1: Analysis of the difference $T_{BB,12}(r_{12}) - T_{BB,12}(r_{12}=2)$. Oceans, JJA 2013, 82°S-82°N.

212	39959	-3.193	3.299	0.022	0.011	0.199	0.123
216	45591	-1.374	3.680	0.004	-0.017	0.199	0.127
220	41081	-1.606	5.351	-0.015	-0.040	0.202	0.123
224	29724	-0.878	1.984	-0.021	-0.044	0.160	0.102
228	10377	-0.644	1.664	-0.018	-0.026	0.124	0.074
232	1084	-0.303	1.087	-0.001	-0.000	0.088	0.040

Table 2: Analysis of the difference $T_{BB,10}(r_{10}) - T_{BB,12}(r_{12})$. Oceans, JJA 2013, 82°S-82°N.

Temperature	Samples	Min	Max	Median	Mean	Standard	Mean
Т _с (К)	Count					deviation	absolute
							deviation
188	73	0.016	0.039	0.032	0.031	0.005	0.004
192	348	0.000	0.221	0.040	0.048	0.034	0.028
196	2149	-0.025	0.345	0.051	0.066	0.050	0.038
200	6915	-0.27	0.667	0.083	0.093	0.060	0.046
204	16783	-0.018	0.729	0.095	0.113	0.078	0.059
208	28620	-0.277	1.155	0.098	0.118	0.086	0.066
212	39959	-1.792	0.763	0.091	0.110	0.085	0.063
216	45591	-0.164	0.601	0.080	0.097	0.071	0.055
220	41081	-0.392	0.622	0.064	0.077	0.058	0.044
224	29724	-0.741	0.768	0.044	0.052	0.041	0.030
228	10377	-0.162	0.251	0.025	0.029	0.025	0.018
232	1084	-0.367	0.093	0.010	0.012	0.026	0.010

Table 3: Analysis of the difference between β_{eff} computed using $T_{BB,10}(r_{10})$ and $T_{BB,12}(r_{12})$ and β_{eff} from this study. Oceans, JJA 2013, 82°S-82°N.

Temperature	Samples	Min	Max	Median	Mean	Standard	Mean
Т _с (К)	Count					deviation	absolute
							deviation
188	73	-0.0001	-0.0000	-0.0000	-0.0000	0.0000	0.0000
192	348	-0.0005	0.0000	-0.0001	-0.0001	0.0001	0.0001
196	2149	-0.0041	0.0001	-0.0001	-0.0002	0.0003	0.0002
200	6915	-0.0048	0.0013	-0.0003	-0.0004	0.0004	0.0003
204	16783	-0.0081	0.0026	-0.0005	-0.0006	0.0006	0.0005
208	28620	-0.0215	0.0080	-0.0006	-0.0008	0.0009	0.0006
212	39959	-0.0220	0.0551	-0.0007	-0.0009	0.001	0.0007
216	45591	-0.0159	0.0075	-0.0007	-0.001	0.001	0.0007
220	41081	-0.0311	0.0129	-0.0007	-0.001	0.001	0.0007
224	29724	-0.0338	0.0319	-0.0006	-0.0008	0.0009	0.0006
228	10377	-0.0109	0.0069	-0.0004	-0.0005	0.0007	0.0004
232	1084	-0.0029	0.0176	-0.0002	-0.0002	0.0008	0.0002

RC S3. Section 2.2.2: What does the bias look like between model and observations before correction?

AC S3. Probability density functions (PDFs) of the differences between clear sky observations and model (BTDoc) at 12.05 μ m before (night: light blue; day: orange) and after (night: navy blue; day: red) correction are shown below for six latitude bands over ocean (left) and over land (right) in January 2008. This figure is now Fig. S1a under Supplementary Materials.



Figure S1a: Probability density functions of the differences between clear sky observations and computations of brightness temperature (BTDoc) at 12.05 µm, before (night: light blue; day: orange) and after (night: navy blue; day: red) correction are shown below for six latitude bands over ocean (left) and over land (right) in January 2008. Note the different scales over ocean and over land.

Similarly, PDFs of the clear sky inter-channel differences [BTDoc(10.6 μ m)– BTDoc(12.05 μ m)] are shown below, and the figure is now Fig. S1b under Supplementary Materials.



Figure S1b: Same as Fig. S1a, but for the inter-channel difference between observations and computations, BTDoc(10.6 µm)-BTDoc(12.05 µm).

The following sentence has been added at the end of Sect. 2.2.2: "Figures S1a and S1b in Supplementary Materials show distributions of BTDoc before and after correction".

RC S4. Section 2.3: You reference Yang (2005) and make some mention of habits on page 10. Are allowing habit to be a free parameter or are discouraging people from using habits at all in prescribing properties from databases such as Yang's? This relevant information to the remote sensing community.

AC S4. We have added a paragraph describing the retrieval's relative insensitivity to ice particle shape at the end of Sect. 2.3. As described in this paragraph (shown below), there is no inferred "ice particle shape recipe" that the retrieval is based on. It would be difficult to apply a database such as Yang's to in situ data since ideally one would need to know the habit composition of the ice particle size distribution (PSD), where habit varies with size across the PSD. Such information is difficult to extract from in situ probe measurements, although it should be possible using the Cloud Particle Imager (CPI) and suitable image analysis software. The new paragraph reads as follows:

"Cirrus cloud emissivity and τ_{abs} depend on ice particle shape (Mitchell et al., 1996; Dubuisson et al. 2008). However, this retrieval should not be very sensitive to ice particle shape for several reasons, one being that θ_{eff} is directly retrieved from cloud radiances as per (2) and (3). Another reason is that no ice particle shape assumptions are made when calculating θ_{eff} from in situ measurements with the exception of the absorption contribution from tunneling (which was not sensitive to realistic shape changes, as described above). That is, the 2D-S probe in situ data include measurements and estimates for ice particle projected area and mass, respectively. MADA optical properties are calculated directly from these in situ area and mass values, thus largely avoiding the need for shape assumptions. Thirdly, this retrieval is most sensitive to the smaller ice particles in a PSD where the variance in ice particle shape is minimal (Baker and Lawson, 2006b; Lawson et al., 2006b; Woods et al., 2018). During the SPARTICUS campaign, many cirrus clouds were sampled so that biases in ice particle shape due to a specific cloud condition are less likely to occur."

RC S5. Figs. 2 and 3 and so on: Why are computed from the TC4 campaign mostly concentrated in regions less than 1.1?

AC S5. As mentioned in Sect. 2.3, PSD sampling times were longer during TC4 (relative to SPARTICUS) with fewer sampling days, resulting in fewer PSD samples. Of these, only aged anvil cirrus sampled on one day by the WB57 were sampled at T < -60°C (Mitchell et al., 2011, JGR). TC4 cirrus sampled at warmer temperatures had substantially broader PSDs that yield lower β_{eff} values, typically < 1.10. New text has been added to the second-to-last paragraph in Sect. 2.3:

"There are much fewer TC4 points in Figs. 2 and 3 for $\beta_{eff} > 1.1$ since the higher β_{eff} values were obtained only for T < -60°C, which only occurred during a single flight."

RC S6. Figs. 8a and 8b: In these figures, IIR median and in situ compare relatively well. Can the authors show an example scatterplot of the comparisons?

AC S6. Because spatially and temporally coincident measurements are very rare for cloudy scenes meeting the IIR cloud selection criteria, these figures show statistical analyses of in situ data on one hand, and of IIR data on the other hand, and are not one-to-one comparisons. Therefore, we don't think that we can show a scatterplot of the comparisons. We added the following sentence at the end of the 1st paragraph in Section 4.1:

"Data analysis is performed on a statistical basis, as coincident in situ and satellite data only provide a very small dataset due to our data selection."

RC S7: Section 5: The selection criteria resulted in less than 2% of qualified pixels. This makes me wonder if the selection requirement is relaxed to include more cirrus clouds, how much do the results in Figs. 11-16 change?

AC S7. The rationale for selecting the relevant cloudy scenes for this study is presented in Sect. 2.2.1. We tried to clarify, and Sect. 2.2.1 now reads (changes are in bold):

"Because IIR is a passive instrument, meaningful retrievals are possible for well identified scenes. This study is restricted to the cases where the atmospheric column contains one cirrus cloud layer. We also insure that the background radiance is only due to the surface (see Eq. (3)) allowing a more accurate computation than for cloudy scenes. The retrievals were applied only to single-layered semi-transparent cirrus clouds that do not fully attenuate the CALIOP laser beam, so that the cloud base is detected by the lidar. The cloud base is in the troposphere and its temperature is required to be colder than -38°C (235 K) to ensure that the cloud is entirely composed of ice. This is likely to exclude liquid-origin cirrus clouds from our data set (Luebke et al., 2016). When the column contains also a dense water cloud, the background radiance can be computed assuming that the water cloud is a blackbody. However, because systematic biases were made evident (Garnier et al., 2012), we chose to discard these cases, which reduces the number of selected samples by about 25 %. Because the relative uncertainties in τ_{abs} and in θ_{eff} increase very rapidly as cloud emissivity decreases (Garnier et al., 2013), the lidar layer-integrated attenuated backscatter (IAB) was chosen greater than 0.01 sr⁻¹ to avoid very large uncertainties at the smallest visible optical depths (ODs). This resulted in an *OD* range of about 0.3 to 3.0. Similarly, clouds for which the radiative contrast R_{BG} - R_{BB} between the surface and the cloud is less than 20 K in brightness temperature units are discarded. IIR observations must be of good quality according to the quality flag reported in the IIR Level 2 product (Vaughan et al., 2017)."

Frequency of occurrence could indeed be increased by relaxing some selection criteria, but the difficulty is that the additional information could be obscured by large additional uncertainties. As an illustration, we reprocessed the dataset to include clouds of OD between 0.1 and 0.3. Fig. S5 under Supplementary Materials show the interrelationships between β_{eff} , α_{ext} , IWC, and N as well as $\Delta N/N$ for OD > 0.1, and is compared to Fig. S4 (former Fig. 7) obtained using the chosen threshold OD > 0.3. Figure S5 is copied below:



Figure S5: Same as Fig. S4 but sample selection criteria was changed to accept samples having OD > 0.1 approximately. Note the larger portion of samples having $\Delta N/N > 1$.

The following text has been added under Supplementary Materials:

"This same analysis was repeated in Fig. S5, except the sample selection criteria for minimum OD was changed from 0.3 to 0.1. This increased the sample population considerably. The larger dispersion in β_{eff} and in particular the larger portion of samples with β_{eff} much smaller than 1 (Fig. S5, top row) are due to large uncertainties at OD between 0.1 and 0.3, which also explain the larger portion of samples with $\Delta N/N > 1$ (Fig. S5, bottom row). More samples now correspond to lower values of α_{ext} (down to 0.016 km⁻¹), IWC, and N. "

And at the end of Sect. 3.2, we added:

"We repeated this analysis except using an OD threshold of 0.1 (instead of 0.3; see also Sect. 5.1). Figure S5 shows this same analysis except that the sample selection criteria for minimum OD was changed from 0.30 to 0.10. In the lower row relating $\Delta N/N$ to β_{eff} , the number of

samples having $\Delta N/N > 1.0$ has substantially increased over both ocean and land relative to Fig. S4 due to the lower OD threshold for sample selection, and more samples correspond to lower values of α_{ext} , IWC, and N."

Table 4 in Sect. 5.1 has been updated to show frequency of occurrence when OD>0.1. The following new text has been added at the end of the 3rd paragraph in Sect. 5.1: *"Relaxing the lower OD threshold to 0.1 instead of 0.3 would increase the number of samples by about a factor of 2.5."*

The equivalent of Fig. 9 (previous Fig. 12a) using a OD threshold of 0.1 instead of 0.3 is now Fig. S10 in the Supplementary Materials. This figure is shown below:



Fig. S10. Same as Fig. 9 in the main paper, but by relaxing the OD threshold to OD > 0.1.

This figure is now introduced at the end of Sect. 5.2 where the following text is added:

"Fig. S10 under Supplementary Materials, shows the same results as in Fig. 9. but by relaxing the OD threshold to OD > 0.1. Consistent with Fig. S5, median N is decreased, by a factor 1.5 on average, while $\Delta N/N$ is more than doubled." A similar statement was added to Summary and Conclusions at the end of the 4th paragraph. As an illustration, but not included in the paper, below are the equivalent of previous Fig. 11 and Figs. 12b to 16 using an OD threshold of 0.1 instead of 0.3.



Same as Fig. 8 (previous Fig. 11), but with OD>0.1. The OD threshold does not change much IIR β_{eff} over ocean. The changes are more notable over land, which could be due in part to larger additional errors than over ocean.



Same as Fig. S9 (previous Fig. 12b), but with OD> 0.1.



Same as Fig. 10 (previous Fig. 13), but with OD>0.1. The relative variations are similar in both figures, with N being smaller when samples of OD between 0.1 and 0.3 are included.



Same as Fig. S11 (previous Fig. 14), but with OD>0.1



Same as Fig. 11 (previous Fig. 15), but with OD>0.1. Consistent with Fig. 8, the changes in median D_e are more notable over land. Larger additional errors occur more over land than over oceans.



Same as Fig. S12 (previous Fig. 16), but with OD>0.1. Again, the changes in median D_e are more notable over land than over ocean.

RC S8. Figure 11. Your highest Beta_Eff occur at the highest clouds where your sample size is often relatively small. Is that an issue? Or is the "hom" effect.

AC S8. It's more likely the "hom effect". That is, we consistently see higher β_{eff} at colder temperatures and higher altitudes with the in situ PSD from field campaigns. This is where hom can produce the highest N, resulting in the smallest crystals that produce the highest β_{eff} . New text has been added at the end of the 3rd paragraph in Sect. 5.2:

"At the coldest temperatures (highest Z_c), hom should be more frequent, resulting in smaller ice crystals and thus higher β_{eff} ."

RC S9. Figure 17: Comparing the two figures, can the authors explain why some peak regions remain similar magnitude such as in southwest Southern America, but many weaken significantly for instance over the Arctic Ocean East of Greenland?

AC S9. It is important to recognize that while the color legend in Fig. 12 (previously Fig. 17) changes color in increments of 50 L^{-1} , the 1st color bin is from zero to 100 L^{-1} and the last color bin is from 500 L^{-1} to infinity. When comparing the two plots, regions corresponding to the highest N values may not change much due to this legend convention.

New text has been added to the end of the 1^{st} paragraph in Sect. 6.4: "While the color legend in Fig. 12 changes color in increments of 50 L⁻¹, the 1^{st} color bin is from zero to 100 L⁻¹ and the last color bin is from 500 L⁻¹ to infinity. When comparing the two plots, regions corresponding to the highest N values may not change much due to this legend convention."