Response to anonymous referee #1

The authors are thankful to the referee for his/her thorough review of the paper.

Our responses are detailed below. The manuscript has been revised accordingly, and is clearly improved.

In the following, the reviewer's comments are in black, and our answer to each comment is in red.

Comments:

1) Regarding references, in the introduction section, an adequate list of references is provided. However, I would suggest the authors to expand the list of references in order to strengthen the manuscript. For example in the very first paragraph, at the end of line 41 (page 2), and at line 42 (page 2) suitable references could be used.

We added the following references at the beginning of the introduction:

Bodas-Salcedo, A., Hill, P. G., Furtado, K., Williams, K. D., Field, P. R., Manners, J. C., Hyder, P. and Kato, S.: Large contribution of supercooled liquid clouds to the solar radiation budget of the Southern Ocean, J. Climate, **29**, 4213-4228, https://doi.org/10.1175/JCLI-D-15-0564.1, 2016.

Muhlbauer, A., McCoy, I. L., and Wood, R.: Climatology of stratocumulus cloud morphologies: microphysical properties and radiative effects, Atmos. Chem. Phys., 14, 6695–6716, https://doi.org/10.5194/acp-14-6695-2014, 2014.

Stephens et al. (2002).

Stephens, G., Winker, D., Pelon, J., Trepte, C., Vane, D., Yuhas, C., L'Ecuyer, T., and Lebsock, M.: CloudSat and CALIPSO within the A-Train: Ten years of actively observing the Earth system, Bull. Amer. Meteor. Soc., 99(3), 569–581, https://doi.org/10.1175/BAMS-D-16-0324.1, 2018.

Duncan, D. I. and Eriksson, P.: An update on global atmospheric ice estimates from satellite observations and reanalyses, Atmos. Chem. Phys., 18, 11205–11219, https://doi.org/10.5194/acp-18-11205-2018, 2018.

Stubenrauch, C. J., Caria, G., Protopapadaki, S. E., and Hemmer, F.: 3D radiative heating of tropical upper tropospheric cloud systems derived from synergistic A-Train observations and machine learning, Atmos. Chem. Phys., 21, 1015–1034, https://doi.org/10.5194/acp-21-1015-2021, 2021.

2) Page 2, line 64: At this point the concept of microphysical index, β_eff., measuring wavelengths, effective absorption optical depths, effective emissivities are introduced in the manuscript. Although the terms are well established, properly explained and presented, this is done later on in the manuscript, leaving a reader to wonder in the early stages of the manuscript. In that case, I would suggest a slight rearrangement, probably would be beneficial for the manuscript, to provide at least brief descriptions at an earlier stage of the manuscript.

We followed the reviewer's suggestion and added a brief description of the emissivity retrievals in this paragraph of the introduction. The text now reads (changes in italic):

"Effective emissivities and microphysical retrievals are reported in the IIR Level 2 data products. The Version 3 (V3) products released in 2011 used the V3 CALIOP data products. As described in G12 and G13, they were focused on retrievals of ice cloud properties. *Effective emissivity represents the fraction of the upward radiation absorbed and re-emitted by the cloud system. The IIR 1-km pixel is assumed to be fully cloudy and the qualifying adjective "effective" refers here to the contribution from scattering. The retrievals are applied to suitable scenes that are identified and characterized by taking advantage of co-located CALIOP retrievals. Effective emissivity in each IIR channel is retrieved after determining the background radiance that would be observed in the absence of the studied cloud system and the blackbody radiance that would be observed if the cloud system were a blackbody source*. Unlike the well-known split-window technique (Inoue, 1985), which relies on the analysis of inter-channel brightness temperature differences, IIR microphysical retrievals use the concept of microphysical index (β_{eff}) proposed by Parol et al. (1991). This concept is applied to the pairs of IIR channels at 12.05 µm and 10.6 µm and at 12.05 µm and 08.65 µm, with β_{eff}12/10 and β_{eff}12/08 defined as, respectively, the 12.05-to-10.6 ratio and the 12.05-to-08.65 ratio of the effective absorption optical depths. The latter are derived from *the* cloud effective emissivities retrieved in each of the three channels. The microphysical indices are interpreted in terms of D_e by using look-up tables (LUTs) built for several ice *habit* models. D_e is retrieved using the ice *habit* model that provides the best agreement with the observations in terms of relationship between β_{eff}12/10 and β_{eff}12/08. Total water path is then estimated using IIR D_e *and visible optical depth estimated from IIR* effective emissivities."

The sentence: "*The IIR 1-km pixel is assumed to be fully cloudy and the qualifying adjective "effective" refers here to the contribution from scattering",* was moved from Sect. 3.1 to the introduction.

3) Page 2, line 69: please provide a more detailed description of the homogeneity criteria used. Although they are detailed in previous studies (Garnier et al., 2012, 2013), as stepping-stone a brief description could be of use.

Thank you for this suggestion. A brief description has been added and the text now reads (changes in italic):

"Retrievals along the CALIOP track are extended to the IIR swath *by assigning to each swath pixel the retrievals in the radiatively most similar track pixel at a maximum distance of 50 km (G12). This most similar track pixel is found by minimizing the mean absolute difference between the brightness temperatures in the three channels, with an upper threshold set to 1 K. Retrievals along the CALIOP track and over the IIR swath* are reported in the IIR Level 2 track and swath data products, respectively."

4) The analysis is mainly in the geographical domain between 60oS and 60oN. Although the biases, the developed algorithms and the improvements are extensively discussed it is not clear the geographical reasons why the analysis is constrained in this domain. I wonder whether the authors can provide an explanation regarding the underlying causes of the geographical preference.

We should have explained that we chose this geographical domain to ensure that the dataset is not contaminated by sea ice (retrievals over surface types other than oceans will be presented in an upcoming paper). This is now explained in Sect. 3.3.1 where Fig. 2 is described (new text in italic):

"The results are shown for 6 months of nighttime data in 2006 (from July through December) between 60° S and 60° N *to ensure that the dataset is not contaminated by sea ice*."

5) Regarding the scene classification, as mentioned, it is based on the characteristics of the layers reported in the CALIOP 5-km cloud and aerosol products. However, as the classification algorithms which is designed to identify suitable scenes containing the required information for the retrievals, sometimes fails to properly classify a cloud/aerosol layer, and moreover in cases of low aerosol/cloud load, due to SNR and

CALIOP detection thresholds/capabilities, may propagate towards the retrievals, the analysis and the uncertainties. It would be beneficial to discuss more extensively in the manuscript the effects of erroneous feature classifications to the retrieval algorithms.
Thank you for this question.

a)The IIR operational algorithm is applied regardless of the confidence in the CALIOP cloud/aerosol classifications. However, we are now explaining in the text that this confidence in the classifications is reported in the product. Cloud/aerosol mis-classifications are expected to be associated to no or low confidence in the feature type classification, and therefore can be easily filtered out using the Quality Assessment flags reported in the IIR product. The impact of a layer mis-classification depends on the contribution of this layer to the measured infrared radiances. In a simple case where the column includes one aerosol layer mis-classified as a cloud, the microphysical retrievals will likely fail because the LUTs are designed for cloud retrievals. In another simple case where the column includes one cloud layer mis-classified as an aerosol, no microphysical retrievals will be provided.

Analyses of both CALIOP and IIR retrievals in case of possible cloud/aerosol mis-classifications are required before we can fully address this question in a satisfactory manner. This would require first to establish why the CALIOP algorithm had no or low confidence in the classifications and then to examine the IIR retrievals to assess whether there is evidence of mis-classification. Such studies could help better characterize the performance of our combined retrievals for these challenging cases.

b) CALIOP detection capabilities are typically superior to the sensitivity required for the IIR algorithm. Thus, the scene classification is based on layers detected by the CALIOP algorithm at 5-km and 20-km horizontal averaging intervals (Vaughan et al., 2009), while layers detected by CALIOP at 80-km horizontal averaging intervals are ignored because they are optically too thin to be seen by IIR.
Nevertheless, CALIOP detection capabilities are limited when a layer fully attenuates the CALIOP signal, so that the lower part of the cloud is missed by CALIOP. For these opaque clouds, the true base is a priori not detected, which introduces uncertainties in the determination of the radiative temperature in ice clouds as discussed in Sect. 3.4.2.

c) Changes in the text:

The text at the beginning of Sect. 2 now reads (changes in italic):
"Both in V4 and in V3, the first task of the IIR algorithm is to classify the pixels in the scenes being viewed. This scene classification is based on the characteristics of the layers reported in the CALIOP 5-km cloud and aerosol products *for layers detected by the CALIOP algorithm at 5-km and 20-km horizontal averaging intervals (Vaughan et al., 2009). This classification* is designed to identify suitable scenes containing the required information for effective emissivity retrievals."

New reference:
Vaughan, M., Powell, K., Kuehn, R., Young, S., Winker, D., Hostetler, C., Hunt, W., Liu, Z., McGill, M., and Getzewich, B.: Fully automated detection of cloud and aerosol layers in the CALIPSO lidar measurements, J. Atmos. Oceanic Technol., 26, 2034–2050, https://doi.org/10.1175/2009JTECHA1228.1, 2009.

And the following text has been added at the end of Sect. 2:
"A lot of other parameters characterizing the scenes are reported in the V4 IIR product. Among them are the number of layers in the cloud system, as well as an Ice Water Flag which informs the user about the phase of the cloud layers included in the system, as assigned by the V4 CALIOP Ice/Water phase algorithm (Avery et al., 2020). A companion Quality Assessment Flag reports the mean confidence in the feature type

(i.e., cloud or aerosol) classification (Liu et al., 2019) and in the phase assignment for these cloud layers. The product also includes the number of tropospheric dust layers and of stratospheric aerosols layers in the column and the mean confidence in the feature type classification. All the suitable scenes are processed regardless of the confidence in the classifications and phase assignments reported in the CALIOP products, so that the user can define customized filtering criteria adapted to specific research objectives."

New reference:

Liu, Z., Kar, J., Zeng, S., Tackett, J., Vaughan, M., Avery, M., Pelon, J., Getzewich, B., Lee, K.-P., Magill, B., Omar, A., Lucker, P., Trepte, C., and Winker, D.: Discriminating between clouds and aerosols in the CALIOP version 4.1 data products, Atmos. Meas. Tech., 12, 703-734, https://doi.org/10.5194/amt-12-703-2019, 2019.

6) Please provide some more information regarding the algorithm performance on thin clouds/cirrus clouds.
In the description of Fig.1, we added in the text that $\varepsilon_{eff,12}$ = 0.1 corresponds to optical depth ~ 0.2, and to a thin cirrus cloud. The text reads (changes in italic):
"at $\varepsilon_{eff,12}$ ~ 0.1 *(or optical depth ~ 0.2, corresponding to a thin cirrus cloud)*, β_{eff}12/10 (dashed line) is decreased…"

7) Is it possible to provide more detailed description on the motivation for changes in V4, through study cases? If the cases are considered to disrupt the flow of the manuscript, I would suggest their inclusion as supplement.
The changes in V4 are motivated by the need to reduce systematic biases in the IIR V3 products that were made evident after statistical analyses of the retrievals. By accumulating a sufficient number of individual retrievals, typically over a month, the random noise could be significantly reduced, but the systematic biases remained. These biases are not unambiguously detected through case studies, because they can be hidden by the noise. We do not think that showing case studies would provide useful additional information.

8) The V4 statistics are very interesting, though they may need further explanation in the manuscript. Is it possible to include in the Statistical Table more statistical indicators (e.g. Relative Difference)?
In this section, we examine the differences between the observed and computed brightness temperatures in order to assess the errors in the computed background radiances used in the effective emissivity retrievals. The relevant indicator is the error in $T_{k,BG}$, rather than the relative error in $T_{k,BG}$, which is why relative differences are not provided in Table 2.

We tried to clarify the text by adding the following sentence at the beginning of Sect. 3.3.2:
"In order to assess the errors in the computed background radiances used in the effective emissivity retrievals ($R_{k,BG}$, see Eq. 1) and in the corresponding computed brightness temperatures ($T_{k,BG}$), we analyzed distributions of BTDoc for different latitudes and seasons."

Furthermore, we modified the text at the end of this section, which now reads (changes in italic):
"*Thus, the analysis of these inter-channel distributions shows that* the uncertainty in computed $T_{k,BG}$ *can be* taken identical in all channels. *Based on the standard deviations in BTDoc(12), the random error ΔT_{BG} is set to the conservative value ± 1 K for all channels*."

9) In 3.4.2 section, I would suggest to include more information on the correction functions, as mentioned briefly in paragraph 2.
We now include the two important equations established in G15, and re-organized the beginning of the section, which now reads (changes in italic):

"As demonstrated in G15, the radiative temperature T_r (k) in channel k is the brightness temperature associated with the centroid radiance of the attenuated infrared emissivity profile within the cloud. For a cloud containing a number, n, of vertical bins, i, of resolution dz, with i = 1 to i = n from base to top, this centroid radiance can be written as a function of radiance $R_k(i)$ of bin i and CALIOP particulate (i.e., cloud) extinction coefficient, $\alpha_{part}(i)$, as:

$$R_k = \frac{\sum_{i=1}^{i=n}\left(1-e^{-\left[\alpha_{part}(i)\cdot\delta z/r\right]}\right)\cdot R_k(i).e^{-\sum_{j=i+1}^{j=n+1}\left[\alpha_{part}(j)\cdot\delta z/r\right]}}{\varepsilon_{eff,k}} \qquad (5)$$

The term $\alpha_{part}(i).\delta z/r$ in Eq. (5) is the absorption optical depth in bin i. The ratio, r, of CALIOP optical depth to IIR absorption optical depth is taken equal to 2 (G15). The radiance $R_k(i)$ is determined from the thermodynamic temperature in bin i.

On the other hand, T_c is the temperature at the centroid altitude of the attenuated 532 nm backscatter coefficient profile, which is written as a function of altitude Z(i) at bin i and $\alpha_{part}(i)$ as:

$$Z_c = \frac{\sum_{i=1}^{i=n} Z(i)\cdot\left(\beta_{part}(i)+\beta_{mol}(i)\right)\cdot e^{-2\sum_{j=i}^{j=n}\left[\eta\alpha_{part}(j)+\alpha_{mol}(j)\right]\cdot\delta z}}{\sum_{i=1}^{i=n}\left(\beta_{part}(i)+\beta_{mol}(i)\right)\cdot e^{-2\sum_{j=i}^{j=n}\left[\eta\alpha_{part}(j)+\alpha_{mol}(j)\right]\cdot\delta z}} \qquad (6)$$

In Eq. (6), $\beta_{part}(i)$ is the cloud particulate backscatter in bin i, $\alpha_{mol}(i)$ and β_{mol} (i) are the molecular extinction coefficient and backscatter, respectively, and η is the ice cloud multiple scattering correction factor (Young et al. (2018) and references therein).

Using V3 CALIOP extinction and *backscatter* profiles in semi-transparent ice clouds, the T_r-T_c difference was found to increase with both cloud optical depth and geometric thickness (G15).

Because the CALIOP extinction profiles are not used in the IIR operational algorithm, the approach in V4 was to establish parameterized correction functions, T_r(k) - T_c, for each channel k, and to correct the initial estimate T_c that was used in V3 as T_r (k) = T_c + [T_r (k)- T_c]. These correction functions were derived off-line from the statistical analysis of a series of simulated extinction and attenuated backscatter profiles. In order to reproduce the variability associated to the various possible shapes of the extinction profiles, we chose to use actual V4 CALIOP profiles (8,000 profiles were used) rather than synthetic profiles. These initial CALIOP profiles were derived from single-layered semi-transparent clouds classified with high confidence as randomly oriented ice (ROI) by the V4 ice/water phase algorithm (Avery et al., 2020). Each CALIOP extinction (and backscatter) profile was scaled to simulate several pre-defined optical depths *corresponding to several pre-defined effective emissivities using r = 2*, and the attenuated backscatter profile was simulated by applying the required attenuation to the simulated total (molecular and particulate) backscatter profile. The simulations *of T_r (k) using Eq. (5) and of T_c using Eq. (6)* were carried out for $\varepsilon_{eff,k}$ ranging between 0.1 (or $\tau_{a,k}$ = 0.1, see Eq. (2)) and 0.99 (or $\tau_{a,k}$ = 4.6). Variations of T_r (k) - T_c with η between 0.5 and 0.8 were also analyzed in order to cover the range of temperature-dependent values used in V4 (G15; Young et al., 2018). Variations with η were not discussed in G15 because η was taken constant and equal to 0.6 in V3.

The $T_r(k) - T_c$ differences were examined against the "thermal thickness" of the clouds; that is, the difference between the temperatures at cloud base (T_{base}) and at cloud top (T_{top}). *Ninety percent of the CALIOP profiles used for this analysis had $T_{base} - T_{top}$ between 10 and 50 K.*"

10) 3.4.3. In the Radiative temperature in liquid water clouds, but also in the rest of the section, I would suggest a more detailed approach and description in the manuscript on the uncertainties introduced due to the applied algorithms. If possible, uncertainties should be included in as many presented results and Figures as possible.

In this section, we present the radiative temperature in ice clouds (3.4.2) and in liquid water clouds.

Radiative temperature in ice clouds (3.4.2)

The correction functions presented in this section were obtained from median values of $(T_r - T_c)/(T_{base} - T_{top})$ as a function of $T_{base} - T_{top}$ based on 8,000 CALIOP profiles. To evaluate the error in the corrections, we added comparisons of the radiative temperatures derived directly using the CALOP extinction profiles to the radiative temperatures derived from the algorithm. The following text and a new Table are now added in Sect. 3.4.2 after Fig. 6:

"The errors in the ice cloud radiative temperature corrections were assessed by comparing T_r derived directly using the CALIOP extinction profiles with T_r derived from Eq. (7). The statistics obtained from the same 8,000 CALIOP profiles as above are provided in Table 3, for both the $T_r - T_c$ correction and the correction error, for channel 12.05 µm. These statistics are provided for $\varepsilon_{eff,12}$ equal to 0.2, 0.6, and 0.99, and using η equal to the extreme values 0.5 and 0.8. The median and mean correction errors are smaller than 0.25 K and significantly smaller than the median and mean corrections, which are found between 0.8 K and 5 K. The standard deviations of the correction errors are between 0.66 and 1.2 K at $\eta = 0.5$ and between 0.7 and 1.75 K at $\eta = 0.8$, while their mean absolute deviations are smaller than 1.25 K. These quantities represent the estimated random error in the cloud radiative temperature correction resulting from the variability in the shape of the extinction profiles.

Table 3: Statistics (median, mean, standard deviation (STD), and mean absolute deviation (MAD)) of the T_r-T_c correction at 12.05 µm and of correction errors for $\varepsilon_{eff,12}$ equal to 0.2, 0.6, and 0.99, using η equal to 0.5 and 0.8. Channel 12.05 µm.

		T_r -T_c correction (K)			Correction error (K)		
		$\varepsilon_{eff,12}$ = 0.2	$\varepsilon_{eff,12}$ = 0.6	$\varepsilon_{eff,12}$ = 0.99	$\varepsilon_{eff,12}$ = 0.2	$\varepsilon_{eff,12}$ = 0.6	$\varepsilon_{eff,12}$ = 0.99
$\eta = 0.5$	Median	0.81	2.04	3.07	0.08	-0.01	0.01
	Mean	0.92	2.22	3.43	0.1	0.02	0.13
	STD	0.55	1.16	1.98	0.66	0.75	1.2
	MAD	0.43	0.93	1.56	0.46	0.53	0.84
$\eta = 0.8$	Median	1.32	3.8	4.50	0.17	-0.01	0.01
	Mean	1.45	4.11	5.01	0.23	0.00	0.17
	STD	0.78	2.08	2.86	0.7	0.98	1.74
	MAD	0.62	1.65	2.26	0.50	0.70	1.24

The error in the radiative temperature estimates is difficult to assess unambiguously, because there is no definite reference to compare the observations with. The errors can be due to the algorithm or to the MERRA-2 temperature profiles.

For ice clouds (Sect. 3.4.2), our main concern is for opaque clouds, because CALIOP sees an apparent base and not the true base. We compared T_r with T_{base} and T_{top} in opaque clouds (Fig. 7), and found that the V4 results are on average in excellent agreement with recent analyses by Stubenrauch et al. (2017), who retrieve a radiative height from AIRS infrared observations. We also compared the measured brightness temperature and the radiative temperature and show that the maximum possible bias is equal to 1.5 K on average.

For water clouds (Sect. 3.4.3), we compared directly the measurements in opaque water clouds with the TOA blackbody temperatures derived from T_r and the radiative transfer model, and we provide with statistics. We also illustrate the impact of the temperature profiles used for the retrievals.