Dear Reviewer and Editor,

here below our answers to your comments (reported for your convenience) are presented as a one-to-one reply and highlighted in yellow. We would like to warmly thank the Reviewer for the accurate revision of our work and in particular for raising one point concerning the discussion of the comparison of our results to L3 satellite products which gave us the possibility to better clarify the objective of the analysis.

Please note that a new version of the Article accounting for the Reviewers comments, is also attached as a reply in the discussion section.

Based on in-situ measurements, this manuscript documents ice and mixed-phase cloud statistics at the Concordia Station in the Antarctic. Various aspects of cloud statistics are discussed. A comparison with satellite L3 data product is also described.

This is a valuable study. Given the scarcity of in-situ measurements and the need to validate satellite retrievals in the polar regions, such study is also much needed by the observational community. Including far-IR, the portion of spectrum rarely used by other observational studies is another shining point of this study. However, some discussions and depictions in the text are not accurate, which I shall describe below in detail. I recommend acceptance after these issues are addressed.

Major comments:

It is well known that cloud fraction statistics from satellites hinges on multiple factors, such as the size of the field of view, the frequency used in the observation, and the detection method (active vs. passive). Besides, passive remote sensing has significant challenges in distinguishing clouds from snowy or icy surfaces over the polar region. It is good to see the authors attempted to compare REFIR-PAD cloud fraction with satellite L3 products (Figure 10) but to thoroughly and correctly interpret this figure is not trivial at all. The discussions related to Fig. 10 ignore a couple of key points: (1) cloud fraction statistics from satellite L3 products are related to footprint size, and none of the L3 products used here has the same field of view as REFIR-PAD; (2) active sensors usually can give more accurate results than passive sensors in terms of cloud occurrence, but their footprint sizes are so different that no way a real “apple-to-apple” comparison can be made.

We are aware of all the caveats highlighted by the Reviewer and for this reason we reported in the text the different sizes of gridded areas when discussing Figure 10. Similarly, the diverse measurements techniques have been mentioned as possible causes of the differences found in the comparison.

In order to make these points clearer the entire discussion of Section 4.3 is revised. The revision involves all the text of the section, and Figure 10 (here below reported for your convenience) is enriched with a right panel which visualizes the extent on the map of the L3 gridded products. Changes are applied to the abstract and the conclusions too.

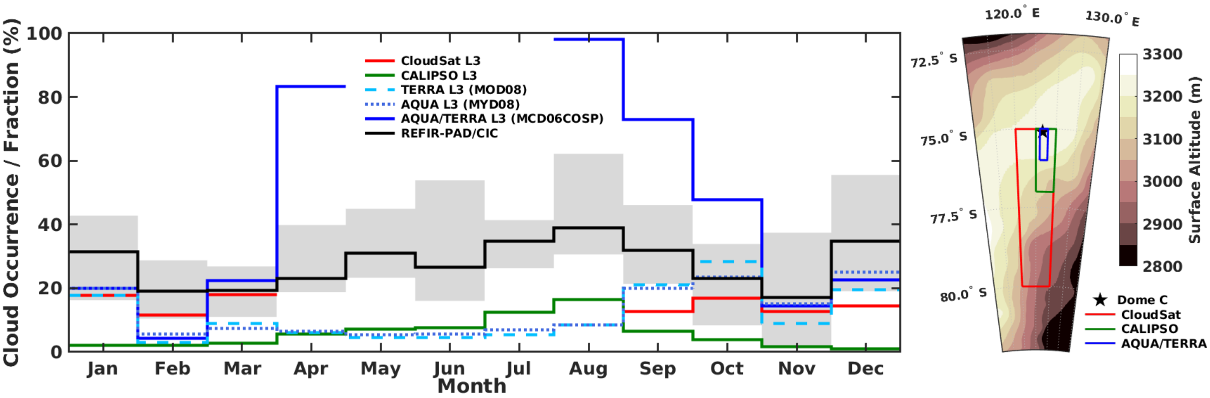

Figure 10.Left panel: Percentage fraction of CIC monthly mean cloud occurrence (in black) compared with CloudSat L3 product (red line), CALIPSO L3 product (green line), and MODIS L3 products (solid blue line for combined AQUA and TERRA L3 product - MCD06COSP, dashed blue line for TERRA L3 product – MOD08, and dotted blue line for AQUA L3 product - MYD08). The shaded grey area indicates the minimum and maximum CIC monthly values in the interval 2012–2015. Right panel: Location of the Dome Concordia base and extension of the grid sector for Cloudsat, CALIPSO and AQUA/TERRA MODIS L3 data. Surface elevation above mean sea level is also reported.

It is also correct that the goal of the comparison has not been sufficiently described. In this regard, we added the following text in the same section (4.3):

"Monthly mean cloud occurrences/fractions derived from level 3 (L3) satellite products are also reported in the left panel of Figure 10 for the same period of time. The comparison has a twofold objective: a) to assess if the results obtained locally from the CIC/REFIR-PAD synergy can be representative of widespread region characterizing the Antarctic Plateau and b) to estimate the differences among the cloud occurrences/fractions derived from L3 satellite products around the Concordia area."

And we also added that:

"Since the L3 products of the three sensors refers to multiple extent areas of observations (of the order of tens of thousands of km2), some differences are expected not only between the ground-based measurements analysed by CIC but also among the mean values of the L3 satellite products. In particular, we note that the gridded L3 products from CALIPSO and CloudSat refer to areas characterized by important variations in surface altitude with possible consequences on cloud formation and occurrence."

And another example of discussion of the results which accounts for the Reviewer comment:

"Nevertheless, higher percentage of cloudiness is found by the CIC algorithm with respect to the CPR. The main reasons for such differences are likely due to: (1) the high CIC sensitivity to the optically thin ice clouds which are often present in the Antarctic Plateau (Maestri et al., 2019) and missed by radar measurements (Henderson et al., 2013; L'Ecuyer et al., 2008), (2) the extension of the gridded area of the CPR L3 product which comprises regions with surface elevations spanning up to 0.4 km in altitude and which might not be representative of the Dome C conditions, and (3) the CPR coarse vertical resolution (0.5 km) which might be the cause of undetected clouds near the surface (Chan and Comiso, 2011)."

An enormous amount of effort has to be invested for data subsetting and collocation in order to make a fair comparison, which cannot be done with the L3 product directly. Figure 10 is useful and informative, but the interpretation here has to be conservative, with all caveats well described before the discussion.

We totally agree with the Reviewer on this point. In fact, this is not the goal of the analysis. We do not aim at providing any satellite products validation in the Concordia region. Such a goal would have required the exploitation of L2 products which are supplied at much higher spatial resolution and consequently can be more accurately collocated. As noted by the Reviewer this operation implies a huge effort on data selection and collocation and it goes beyond the scopes of the present paper. Anyway, we think that the plot provides insights on the accuracy of information that can be extracted from L3 datasets. The L3 products are easily accessible and provides information on the average atmospheric conditions (such as monthly mean cloud occurrence) of gridded area. When the same kind of information becomes available for a specific location independently of the satellite product, a comparison with the independent data is always instructive about the ability of gridded data to characterize a specific location.

Text is added in Section 4.3 in this regard:

"According to WMO[1], the L3 satellite products are composed of variables mapped on uniform space-time grid scales and are constructed to provide completeness and consistency for the anticipated users. These products types are frequently used to perform climate analysis and model evaluation (e.g. Stubenrauch et al., 2013; Webb et al., 2017). The assessment of their accuracy can be particularly challenging, especially in remote regions such as the Antarctic Plateau, due to the scarceness of ground-based stations that are available for products validation campaigns. For the present study, we only refer to monthly mean L3 satellite products and the comparison with CIC results is performed only in the context of the objectives described above. A validation (that is outside the scopes of the present research) should be, eventually, performed on level 2 collocated satellite products to minimize the bias due to different footprint sizes that can be otherwise very large when accounting for gridded L3 products. In practice, different data sets present specific strengths and limitations that are briefly described below."

For example, the abstract stated, "A comparison of monthly mean 15 results with cloud occurrences/fractions derived from level 3 satellite products, from passive and active sensors, emphasizes the difficulties of satellite observations in the Antarctic region and highlights the ability of the CIC/REFIR- PAD synergy to identify multiple cloud conditions and studying their variability at different time scales.", which I think is not a fair statement given all the reasons mentioned above.

True. The abstract is modified as follows: "Monthly mean results are compared to cloud occurrence/fraction derived from gridded (Level-3) satellite products, from both passive and active sensors. The differences observed among the considered products and the CIC results are analysed in terms of footprint sizes and sensors' sensitivities to cloud optical and geometrical features. The comparison highlights the ability of the CIC/REFIR-PAD synergy to identify multiple cloud conditions and study their variability at different time scales."

Other comments:

Figure 2c, red spectra, I am surprised to see the clear-sky spectrum here has a peak at CO2 band center (~667 cm-1) as low as 150 K BT. Since this is a surface measurement looking up, the BT at the CO2 band center should be close to the temperature in the lower troposphere or even in the boundary layer. Thus, it cannot be so low. The cloudy spectra here, as well as the clear-sky spectra in Figure 5, look all reasonable to me. Thus, this 150K BT peak at the CO2 band center in Figure 2c needs to be examined and explained.

The brightness temperature peak at 15-um is a calibration artefact due to the noise amplification (due to calibration process) present at this wavenumber caused by the strong air absorption inside the interferometric path. Since the noise is not shown in the figure, for a better clarity, we will remove this spectral point in the revised figures (both figure 2 and 4). The noise plots of REFIR-PAD measured in the Antarctic campaign are reported in https://doi.org/10.5194/amt-12-619-2019.

Please note that radiances at wavenumber from 620 to 670 cm-1 (in the CO2 band, with the inclusion of the Q-branch at band center 667 cm-1) are not ingested by the CIC algorithm and thus the indicated bad calibration channel does not affect the classification results. This is also the case of others bad calibrations points at FIR channels below 300 cm-1 and above 1200 cm-1, caused, respectively, by strong absorption from H_2O rotational lines, and by the absorption of the Mylar beam-splitter. These bands too are not ingested by CIC which is applied in the 380-620 and 670-1000 cm^{-1} spectral intervals.

The information about the exclusion of the 620-670 cm-1 spectral interval was provided in previous papers such as the one describing the CIC algorithm [Maestri et al., 2019] and the one accounting for its first application to airborne data [Magurno et al., 2020]. Since the information is missing in the present paper, a new sentence is added for completeness:

"Note that, as discussed in Maestri et al. (2019b) and Magurno et al. (2020), the spectral interval 620-670 cm−1 is excluded by the analysis."

Line 130: The impact of multiple scattering within liquid clouds on the depolarization ratio can be included as justification for the 15% depolarization threshold.

Yes. We added the following comments: "In this study a depolarization ratio of 0.15 is used as a threshold for the discrimination of the liquid water clouds and ice clouds over the Concordia Station. The value accounts for possible increases due to multiple scattering effects as discussed below."

and

"The 15% depolarization ratio value is selected to account for the impact of multiplescattering within liquid clouds. It is observed that in presence of mixed-phase clouds the depolarization ratio shows very small values at cloud base, characteristics of liquid spheres, and increases towards values typical of ice crystals near the cloud top. An increase is, in part, intrinsically related with liquid water layers, where multiple scattering determines a depolarization that gradually increases with the depth of penetration, in the lidar backscatter. For this reason, in some conditions, the phase of the upper part of the cloud cannot be unambiguously defined based on the analysis of the depolarization ratio profile only. Nevertheless, the presence of liquid phase at bottom is unequivocally identified and the cloud is categorized as mixed-phase."

Line 154: These classes are not explicitly mentioned in the abstract, please list them there.

Now the abstract reports:
"The CIC algorithm is optimized for Antarctic sky conditions and results in a total hit rate of almost 0.98, where 1.0 is a perfect score, for the identification of the clear sky, ice and mixed-phase clouds classes."

Table 3: Misclassifications are in the hit rate column. Please make it clearer that these are misclassifications through labeling or reformatting the table.
A new column is added to Table 3 accounting for the misclassified data

Figure 7: The caption mentions unclassified spectra, but there does not seem to be any in the figure.

True. An automatic software for plotting was used which accounts for all the possibility. A new figure is generated with a correct caption.

Line 345: The authors mention that a subset of the long-term data is used for training and testing. Is this different from the training data and testing data mentioned in the preceding text? If so, please state. If not, then please indicate why the algorithm is retrained and reoptimized.

Rephrased:
"A total of 87960 REFIR-PAD spectra are analysed from the dataset spanning over the time range 2012-2015. From this set, only 202 spectra are used for training the CIC algorithm, and the other 87758 are ingested by the CIC to evaluate the cloud occurrence over the Concordia station."