First of all we sincerely thank the reviewers for their work in reviewing the manuscript. These reviews were helpul to improve the manuscript and we hope that their justified concerns will be answered here. The reviewer's comment is in normal color, while the answer is in blue.

Reviewer 1

This is a thorough and important study demonstrating how the CloudNet radar network can be leveraged to provide intercalibration between ground-based and spaceborne cloud radars; it expands and improves upon the previous work by Protat et al. and Kollias et al using similar methods, and will provide an important means of monitoring and calibrating ground-based radars and EarthCARE during its lifetime. The manuscript is well-structured and clearly written, and the figures are well-presented; however, there are some methodological issues which need addressing to make the treatment of EarthCARE data consistent with the ground-based and CloudSat data. I recommend this paper for publication subject to major revisions.

1. For the selection of ice clouds from the ground-based and CloudSat datasets, the CloudNet and DARDAR-MASK target classification products are used to identify liquid cloud layers. Both of these products are synergistic radar-lidar target classifications, wherein the lidar provides the bulk of the detections of liquid cloud. For the current analysis of EarthCARE CPR data, however, it appears that the CPR L2a product (C-TC) is being used. While C-TC contains a classification of liquid cloud (alongside many detections of rain and melting layers, as described in Section 4.3), it will miss the vast majority of non-precipitating liquid cloud—and especially supercooled liquid cloud. This was shown in Irbah et al. (2023). In fact, the C-TC product is not capable of diagnosing supercooled liquid at all: all hydrometeors detected at temperatures below the freezing level will be classed as ice or snow. It may be that the existing criteria for selection of data are screening these cases out to some extent, but there's a good chance that mixed-phase clouds are not entirely removed from the EarthCARE data at present.

We understand this concern as it was also one of ours during this study. The AC-TC classification was not used for several reasons. The first one being known issues in the classification for the BA baseline. An issue corrected in BC baseline was the missclasification of liquid clouds to aerosols (https://earth.esa.int/eogateway/documents/d/earthcare-data-handbook/earthcare-ac-tc-level-2b-processor-disclaimer). We therefore prioritize the removal of liquid clouds over the attenuation from supercooled water. Looking at our final closure, this tradeoff does not seem detrimental as we have a very good agreement between EarthCARE and the ground stations.

Another concern we have is that as supercooled water is detected through the lidar, the geometry of the observation could be a problem. As EarthCARE observes from the top of the cloud, a supercooled water layer embedded in the cloud could be invisible if the lidar is attenuated before it. The same inverted phenomena could happen from the ground. An instrument seeing supercooled water could therefore lead to not removing it for the other, thus keeping an observation of a cloud in the dataset that is not in the other. This would need to be investigated if we want to precisely filter the profiles with supercooled water.

Figure R1 is the equivalent of figure B1 from the manuscript, with supercooled water filtered using AC-TC for the satellite and CloudNet classification for the ground. Removing the supercooled water with AC-TC divides by 2 the number of available satellite profiles and the infered bias is larger. Taking account of the uncertainties, the results with and without

supercooled water are coherent one with another. Looking at R1, we also see that between 8 and 10 km, the ground and satellite center profiles are similar, but this similarity fades at heights lower than 8km.

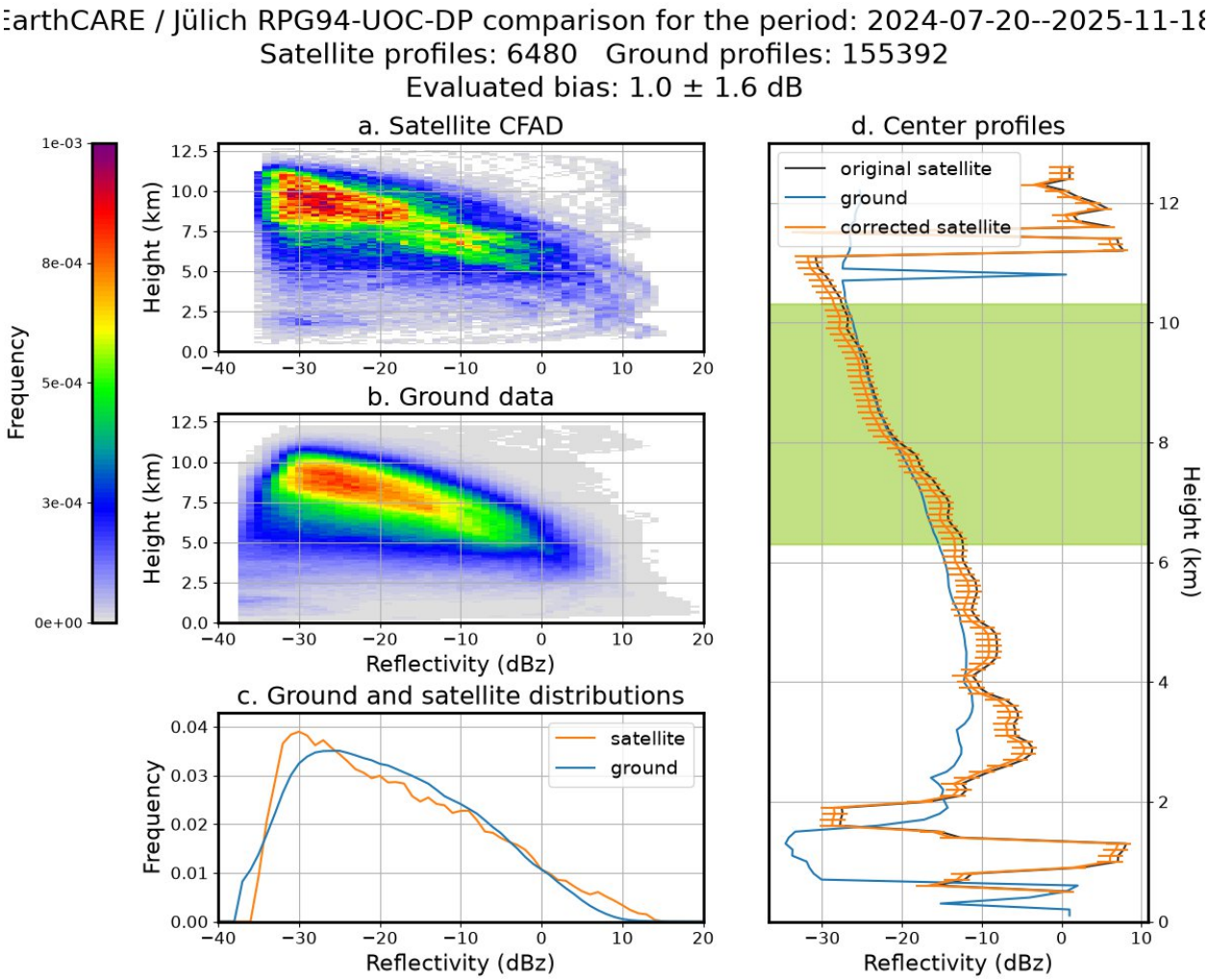

Figure R1: Same figure than figure B1 from the manuscript but with supercooled water filtered using AC-TC. a., b. Satellite and ground CFADs. b. Satellite (orange) and ground (blue) reflectivity distributions. d. Center profiles of the ground (blue), satellite (black) and satellite corrected for the bias (orange).

To summarize, the removal of supercooled water should be possible and could be beneficial when the classifications products have undergone more validation. We think that this filtering would need further investigations and should be done carefully. For our methodology, this could also require to refine our height selection parameters.

Changes to the manuscript: Added a discussion on why we did not used AC-TC, lines 128-130. Added a discussion on the future use of the AC-TC product to filter supercooled water, lines 474-476 .

2. Another apparent issue in the data screening is indicated by the presence of a local maximum in the EarthCARE CFAD (e.g. Figure 3) around -30 dBZ and 2.5km above ground level. Although I see that there is a (smaller & shallower) maximum in the

ground-based data near here, I suspect there's a contribution to the EarthCARE CPR detections from a measurement artefact. This feature is highly characteristic of a known second-trip echo effect that occurs over highly reflective surfaces (including rivers, lakes, ponds, flooded fields, as well as salt-flats and the open ocean in light wind conditions). It was not yet removed from the L2a CPR products at the time of the BA baseline, but will be identified & removed in later versions. These second-trip echoes will occur only in profiles where the surface return (sigma0) in the CPR data exceeds around 30 dBZ, so could be screened for using this criterion.

In the case of our study, these second trip echoes are not a problem and can even be considered as a proof of concept. These heights are indeed rejected for the comparisons by all 3 of out selection criterias, confirming our capability to discard non comparable heights.

Changes to the manuscript: The presence of these scond trip echoes is mentionned in the description of fig3. The confirmation that the algorithm is capable of discarding different physical observations such as scond trip echoes is also mentionned lines 319-322.

3. The only other comment I have is very minor: In the CFAD plot introduced in Fig 3., the "discarded data" section of the ground-based CFAD corresponds to the minimum detectable signal of the EarthCARE CPR. But in some of the plots in the appendix, the lower thresholds of the CFADs and PDFs of reflectivity (panel c) are not consistent between the ground and satellite: Figs. A1, B4, B5 & B6.

Answer: The windowing of the data is made dynamicaly, based on the reflectivity target observed by both instruments. This is done to avoid errors in the calibration bias estimation that could arise when removing data using fixed reflectivity thresholds. A fixed threshold would be quantitatively incorrect if the 2 radars are not calibrated the same way. The window is infered from the satellite PDF then applied on the ground data. The reference to apply the window is the 20% of data from the maximum of the pdf. This 20% might differ between the gound and the satellite, leading to a different minimum reflectivity, thus explaining the difference of lower threshold between EarthCARE and the ground.

Changes to the manuscript: A mention of this potential non coincidence of the minimum of the distributions was added lines 244-245.

Reviewer 2

1. Instead of introducing a fixed time-window around overpass time, they just used all of the ground-based radar data obtained during the comparison period of 15 months. This implies that CPR detected clouds are different from ground-based radar detected clouds. However, there is an assumption that CPR detected many clouds within 200km around an ACTRIS site contain the same ensemble of clouds detected by ground-based radar in the site although different time is chosen. The assumption is not clearly written/defined and justification of ignoring the time issue might be insufficient, e.g., diurnal variations of cloud structure and nature might occur. Some discussion of this point might be given. If they will add above mentioned points and related discussions, the manuscript will be hopefully improved.

This concern was also one we had during this study. Concerning the diurnal variations of the cloud structures, all the sites used in this study have day and night overpasses of the same order. Figure R1 shows an exemple of the contribution from both day and night overpasses for one of our site. We see that day and night overpasses have an equivalent contribution, this is also observed for the other sites of our study.

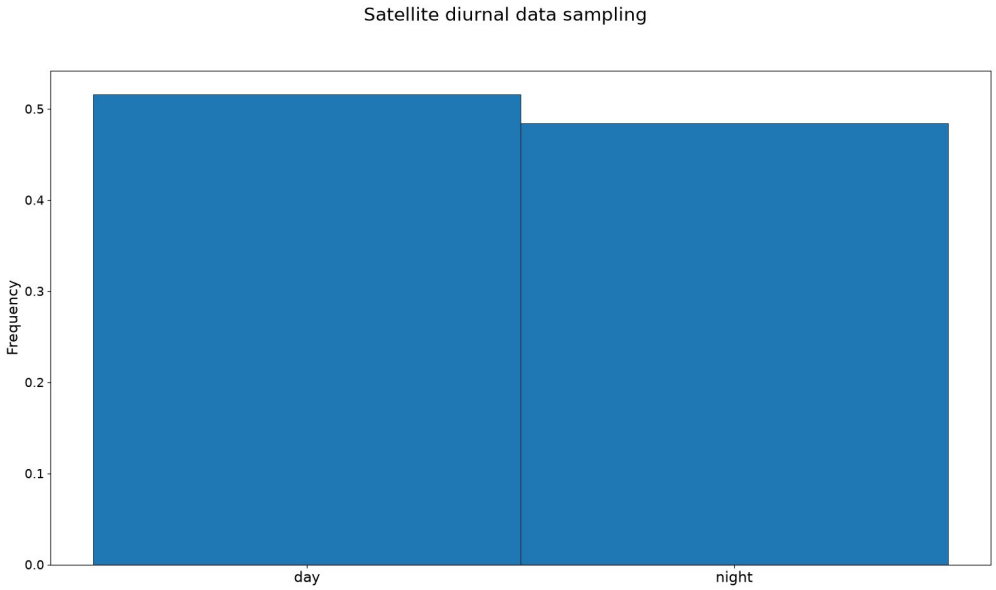

Figure R1: Frequency of day and night overpasses for the site of Juelich on the period from 2024-07-20 to the 2025-11-18.

We also had a look at the monthly data contribution. This is shown on figure R2. From this figure, we see that even though there are variations, the contribution is evenly distributed on the time period considered. The smaller contributions of the ground data during the first few months is due to the fact that the radar was switched off on some time intervals between July and the beginning of December 2024.

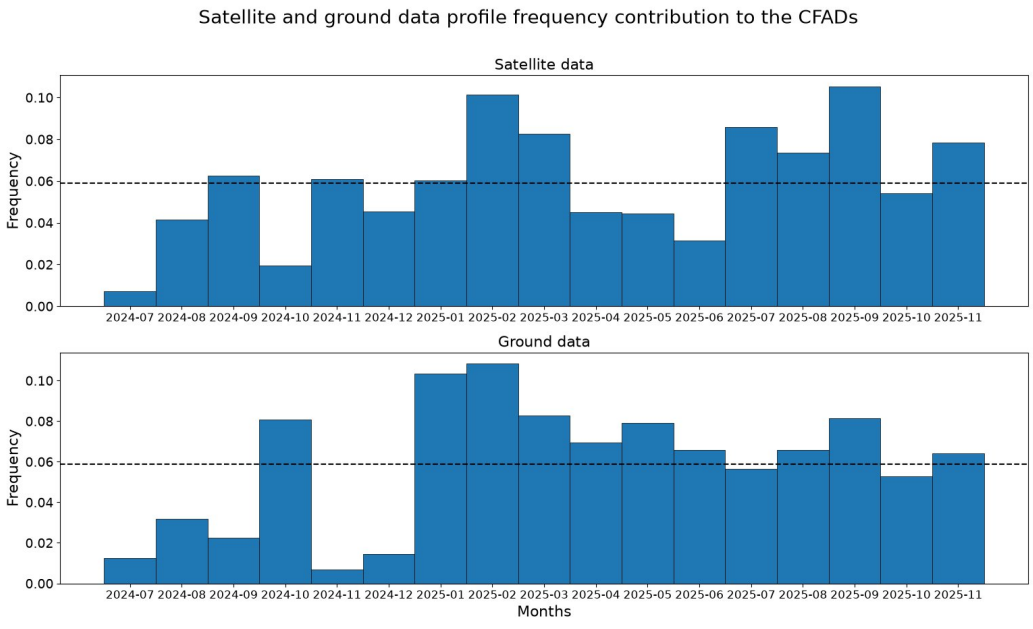

Figure R2: Frequency of the contribution to the dataset from the satellite (top plot) and ground (bottom plot) for each month for the site of Juelich on the period from 2024-07-20 to the 2025-11-18.

The decision to use all of the ground data instead of only data in a ±1h window around the overpass comes from the observation that in the majority of cases, the ground and satellite observe different clouds. On the 183 overpasses in the 15 months considered, less than 20 show similar clouds, hence why a long statistical comparison is needed. Single overpasses figures for a site will be added in supplementary material to show this. Limiting the the collection of ground data in a ±1h window drastically lower the number of profiles from the ground, as shown in appendix A. We therefore consider that evaluating the bias at the most statistically similar height bins leads to comparing similar clouds observed by both the ground and satellite. The weight of eventual outlayer data introduced by considering all the ground profiles should also be reduced by the fact that we consider the center of the fit to evaluate the bias, thus reducing the weight of outlayer data. As we see from Appendix A, considering a ±1h window or the whole ground data leads to similar bias results with our methodology. This leads us to think that with enough time, both ground data selection methods should converge to the same bias result.

Changes to the manuscript: As the discussion on this is in Appendix A, it was modified to improve its relevance. Supplementary material with figures from all the overpasses for the site of Juelich will be added.

2. Product name and the baseline(version) of the product should be explicitly written.

All mentions of products and baselines were checked. Lines 108, 124, 125, 126, 129, 211, 341, 371, 376

Changes to the manuscript: Lines 108, 125, 341, 371, 376, 444, changed to precise CPR FMR product. Line 435 changed to precise CPR TC.

3. In page 10-11, they wrote minimum detectable radar reflectivity is -35dBZ but in Figure 3, data with lower than -35dBZ seem to be used for comparisons. If they actually used data below -35 dBZ, the value of minimum Ze should be written.

We indeed had a binning issue in the figures where all the data was shifted by 1dB. This shift is not physical and only due to the figure configuration.

Changes to the manuscript: All figures with a CFAD (Figures: 3, A1, A2 , B1, B2, B3; B4, B5, B6, B7) were redone to correct this issue.

4. In page 11, Figure 3 2.a and 2.b, shapes of the frequency of reflectivity are different between CPR and ground-based radar. More fraction of small Ze are detected by CPR than by the ground-based radar. Contrary, in p.13, Figure 4 shows the shape are more or less similar and center position of the distribution is similar between the two. Clarification is needed.

Figure 3 shows the reflectivity frequency of the whole datasets, while figure 4 shows the reflectivity frequency for the 8.6km height bin for both satellite and ground. This shows that although the global reflectivity frequencies are not very similar, considering each altitude independently shows more similarities between the datasets. This can be read lines 283-284 of the manuscript. Please let us know if you think clarification is needed.

Changes to the manuscript: Figure 4 was redone with bigger font to ease the distinction.

5. In page 13, line 287, the first parameter. It seems only two values of centers to derive correlation. Then it is not clear how to derive the correlation of two centers.

For each height bin, the correlation between the ground and satellite centers is given by the correlation between all the points in a ±500m vertical height window around the considered height bin. This gives 11 points to evaluate the correlation. This is precised line 291-292 of the manuscript. Please let us know if you think clarifications are needed.

6. In Page 15, height bin selections relied on three parameters. For example, sigma ratio is lower than 0.4. Please add some discussions why the three values are selected.

The choice of the parameters threshold values were made with the following considerations: For the centers correlations, the selection value should be the highest as possible (a high correlation meaning a high vertical similarity between the reflecivity). For the sigma ratio, the selection value should be the lowest as possible (lower sigma indicates a higher statistical similarity at the given height bin). For the R^2, the selection value should be the highest as possible to ensure that the fits accurately represent the data. The three values were then chosen empirically with these considerations in mind while trying to keep the highest number of selected height bins at the end of the selection process. The sensitivity test we performed showed that tunning these thresholds (corr: ±5% R^2:± 5%, sigma: ±25%) does not change the final bias. However these parameters can be adapted if the situation requires it, for example if there is not enough data available for convergence of the bias estimation.

Changes to the manuscript: A discussion on the choice of the parameters values was added lines 303-309.

7. In page 18, line 385, Since same minimum Ze were used for both CloudSat and the ground radar to derive biases CloudSat, is it correct that the bias change attributed to the change of power of CloudSat?

We do not attribute the change of bias to CloudSat but to the ground radar. As CloudSat calibration has been very stable through time (https://airdrive.eventsair.com/eventsairwesteuprod/production-nikal-public/7b9ae858c1db40028ec1c6e6144a3d4d), we use CloudSat to evaluate the calibration of the ground radar as it was done in previous studies (Protat et al.2009, 2011 and Kollias et al. 2019). Moreover the variations observed with the CloudSat are the same than the ones observed by ground teams of the Lindenberg site, comparing the MIRA observations to an MMR

at the site. The use of CloudSat serves a dual purpose. First it shows that timeseries of comparisons can be done with our algorithm. Second, it allows us to study the comparison time needed to obtain a reliable bias as we can have long time series using CloudSat.

Changes to the manuscript: The discussion lines 393-398 was added to be more precise on which instrument causes these bias variations.

8. In pages 23 and 24 Figure A1 and A2, dynamic range of color bar for frequency in A1 is different from A2. Please use the same. Same for B3 and B4.

Our figures do not have the colorbar for the frequency as they don't have the same number of data. If we would use the same colorbars we would loose some contrast in the CFADs for cases with a larger number of data.

Community comment 1

This comment refers to Section 6, Figure 7 of the paper. The variations in the reflectivity bias cannot be attributed to the installation of the clutter fence, as it does not obstruct the radar beam; therefore, no attenuation caused by the fence is expected. Following consultation with the radar manufacturer, the variations in the reflectivity bias are considered to be at least partly attributable to a misalignment of the phase correction position (PCP) in the signal processing chain of the magnetron-based pulse-Doppler radar MIRA. Due to the random phase of the transmit pulses of the magnetron transmitter, the phase reference must be determined at a well-defined position within the transmitted pulse. Contrary to initial assumptions, the PCP was not automatically adjusted or tracked by the system so it had to be set manually.
If this position is not optimally configured, or if it drifts over time (e.g., due to aging of the modulator tube), the phase correction becomes suboptimal. This leads to a reduction in pulse-to-pulse coherence and consequently to a loss in effective sensitivity, which can manifest as a systematic underestimation of radar reflectivity.
Since 2019, the PCP has been regularly monitored and manually adjusted to account for such drifts, which has improved the stability of the reflectivity measurements.

Thanks Ulrich for your nice feedback, helping us to interpret the results from this figure. We added a special mention in the acknowledgmeny for your contribution.

Changes to the manuscript: The discussion line 398-403 was changed to account of these informations. Acknowledgements changed, lines 545-546.

Community comment 2

Dear Nathan, dear all,

First of all, thanks a lot for making this nice and important study available to the community!

1. Purpose of this short comment is just to address the aspect of the community engagement in the acquisition of the underlying datasets. Many of the measurements used in your study were only made available by special funds. Specifically, ATMO-

ACCESS should be mentioned here. E.g., "The preparatory work for EarthCARE validation has been supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, through the ATMO-ACCESS Integrating Activity under grant agreement No 101008004." But also the ACTRIS funds provided by the individual member countries are very relevant, as, e.g., TROPOS/Leipzig would not be able to run any 94GHz radar without the ACTRIS-D funds. E.g., for TROPOS: "ACTRIS-D is funded by the German Federal Ministry of Research, Technology and Space (BMFTR, formerly the German Federal Ministry of Education and Research (BMBF), grant no. 01LK2001A) under the FONA Strategy “Research for Sustainability”. See, e.g., the more community-driven approach of the EarthCARE ATLID Cal/Val study of Baars et al. (2026): https://doi.org/10.5194/egusphere-2026-1490. I also see other ACTRIS funding numbers missing, e.g., for the Basta calibration campaigns.

Overall, the investments don't get visible in your study, even though they were the prerequisites for the used datasets (and partly provided solely for the sake of EarthCARE CAl/Val). Unfortunately, they also don't become visible by just pointing to the Cloudnet database (as it is done in the discussion manuscript). Here, I further suggest to use the option of the Cloudnet web portal to create individual DOIs for each dataset you were using. The CLU Team can support with that, or you just select a site & date range --> select "download" --> "create DOI".

I would appreciate if you could extend the acknowledgements accordingly, so that the activities underlying your study can be tracked appropriately.

Thanks for pointing out this issue. We had a discussion with the ACTRIS CCRES head office to promptly modify the acknowledgments. We also generated a doi for both the ground and satellite dataset.

Changes to the manuscript: Acknowledgments were modified. Added the DOI of the datasets in the data availability.

2. A small final additional comment (while I leave the actual review to the reviewers): It would be nice to list the date range of evaluation already in the abstract (and not just in Tab. 1). This might become a relevant aspect with increasing temporal extent of the EarthCARE dataset.

Changes to the manuscript: The date range of the study was added to the abstract. Line 9