

Review of: "EarthCARE Commissioning Calibration and Validation Campaign in Ottawa (ECALOT): An Integrated Approach to Multi-level EarthCARE Validation" by Qu et al.

General remarks

The manuscript presents an overview of the ECALOT airborne campaign and associated ground-based observations that were designed to provide observations for validating measurements and products of the EarthCARE satellite. ECALOT started in late 2024 and thus is one of the valuable validation missions conducted in the commissioning phase of EarthCARE. The airborne observations focus on a set of remote sensing instruments comparable to those on EarthCARE. In addition, in situ sampling of cloud particles and standard meteorological parameters were obtained. This combination allows a validation of EarthCARE level 1 and level 2 products. Including the ground-based observations, one peculiarity of ECALOT is the potential to validate the EarthCARE off-track observations and the scene reconstruction for those viewing angles. In this perspective, the presentation of ECALOT, its specific design of the research flights allowing center track and off-track validation, is a highly important step in order to promote the data and its use for in-depth validation of EarthCARE products and in addition also numerical weather models.

However, in my view, the manuscript in its current state partly fails to promote the data set. The authors mainly focus on the general approach, flight design and methods for EarthCARE validation and then show some specific examples of comparisons. There is not much to criticize on the data description and the presentation of the validation approach, but there are two major issues that significantly reduce the impact of the manuscript. First, the structure of the manuscript is poor with a lot of duplications. Often general statements that hold for any other satellite validation study but are not specific for ECALOT are given. This should be improved to make reading easier and faster. The second issue I see is the final part where the authors try to demonstrate the potential of the ECALOT data set for validation by some examples. This demonstration is not convincing and does not provide solid conclusions or even suggestions how to improve EarthCARE retrieval products. I understand that this is a typical approach of similar campaign overview papers, but in this case it just does not help and reads half-baked. I hope the authors agree that these issues need to be addressed before publication. Below you find some more detail and suggestions.

Major comments

Structure, duplications, sharpening

To my taste, the manuscript structure can be improved to sharpen the main messages and ECALOT specific aspects of EarthCARE validation. The manuscript often reads like a white paper explaining the very general objectives of ECALOT. And often these general thoughts are repeated throughout the manuscript. I see the potential of the manuscript in a more scientific overview of the ECALOT data set showing the potential of the data and also how it distinguishes from other validation data. How everything worked out (in respect to the strategy, not in respect with the quality of EarthCARE data)? How valuable is the ECALOT data for validation? And prove this by an analysis of the data. These aspects are sometimes hidden in the woven structure of the manuscript and can be combined more efficiently.

The following comments address parts of this issue. Some specific comments listed further below may also help.

Section 2.3 At this point I still do not know about the instrumentation details of ECALOT. Would be helpful to provide this information before claiming that you can validate all these EarthCARE products.

Section 3.1 This section is quite general and likely everything can be read in Amiridis et al. (2025). What would be more important to read is how ECALOT deals with all the specific requirements. This section leaves this open. So I do not see a value repeating Amiridis et al. (2025) without putting it into the context of ECALOT.

Section 3.2 A lot of statements in section 3.2 are repeated and duplicated without getting to the point (e.g., **L 237-238, L 242-243**. In Section 4 comes another section "observation strategy". This can be shortened. Following my comment on section 2.3, I suggest to merge section 2.3 and 3.2.

Section 4.2 A small map of the locations could help to support this section? Readers might not be familiar with the locations and can not quickly think in Lon/Lat. Or is the distance of all locations not relevant to the typical distance of EarthCARE overpass tracks? Figure 3 that is shown later would fit here pretty well. This requires just some reordering of the sections. I can imagine, that Section 4.4 incl. Figure 3 can easily be shown earlier.

L 356-358 Duplication of lines 338+.

L 372-375 Again duplications and very general statements that do not help the reader to understand the specifics of the ECALOT data set.

Tables 1, 5, and 7 The three tables 1, 5 and 7 provide duplicated information. Table 1 and 5 both show measured quantities. Table 5 and 7 both address EarthCARE products to be validated. This information can be comprised by merging the tables. Stand alone, the tables also only show little information. For example, Table 7 is just a list of EarthCARE products that does not convince me, that these products can be validated. This has to be linked to the ECALOT data and an idea what instrument can be used for validating each product.

Section 5.2.1. Section 4 was already on the ECALOT instrumentation. The question is if these sections can be merged? Or at least, I recommend to state in section 4, that details on instrument and data products will be given in section 5.

L 501-503 This has been discussed already in section 4.3 and does not need to be repeated in such a general way here.

Section 6

This section seems not to follow a specific focus. Often general statements on EarthCARE validation strategies are repeated (duplications again). Sometimes data examples from ECALOT are presented. These examples are teasers for what could be done for EarthCARE validation. However, the rather basic and often not quantitative analysis shown here does not allow any conclusion for the EarthCARE validation (see also specific comments). Thus, the scientific value of this section is questionable. On the other side, I would agree, that a substantial validation of EarthCARE products should also not be the focus of this manuscript. This is for more specific in depth analysis paper.

So instead of teasing the EarthCARE validation with only brief and non-trustworthy analysis, I suggest to focus on the ECALOT data in this manuscript. Remove the validation attempts. Instead, show only ECALOT data without comparison with EarthCARE. There is sufficient important content when you demonstrate what data, chases, cloud types are included in the data base. Such an overview with some statistics and a stronger characterization of the data can do it. E.g. a statistical analysis of the data set illustrating the range of conditions covered by ECALOT would help more than presenting single cases. E.g.: What altitude and what temperatures in situ cloud particles were sampled? What were the typical cloud top altitudes? How many clouds were precipitating,...? Similarly, a quantification of the success of flight strategy can be shown. How close in time distance are the observations to the EarthCARE tracks? How well the flight

strategy accounting for wind drift worked out?

You will likely come up with other questions and statistics that demonstrate the character of the data set.

List of specific comments

entire manuscript: Explain all acronyms when first used. This begins with EarthCARE in the abstract.

L 86: For ARCSIX you may cite the data set description paper although still in review/discussion. (<https://essd.copernicus.org/preprints/essd-2026-241/>)

L 85-93: I suggest to add a footnote explaining all the campaign acronyms. I don't think the table in the appendix is helpful. It only provides the acronyms as additional information. References as given in the text are missing in the table anyway. A footnote should do it and makes it easier for readers.

L 94: Is ECALOT filling any gap of EarthCARE validation that is not covered by all the other campaigns? If so I recommend to highlight this here.

L 131: Do you need all these references? You wrote "e.g.". Same for **L 159:**

Table 1: When available I recommend to add references explaining the technical details, performance and uncertainties of these instruments. This partly comes in the application sections. But it is more logical and easier to follow when references are given here. Same holds for Table 2.

Table 1: For HiSRAM) provide the frequency range.

Section 4.3 This section has some more depths but still lacks of quantifications and ECALOT specific conclusions. E.g.: **L 294-298** This are rather general statements and common knowledge. Could you quantify the impact of temporal effects in you data?

L 294-298 Is wind speed and advection of cloud systems considered in this flight planning? 10 m s^{-1} wind velocity would cause about 9 km advection in 15 min flight time. This advection can be in either direction depending on wind direction and may help you to keep on track or help you to stay away from the EarthCARE track.

Figure 1: What does it mean, that the distance of the surface validation should be < 90 km?

Figure 2: "exploration" "underpass" and "full profiling" leg are not discussed in the text. How did you plan the separation of these legs taking into account the wind direction? Can you prove by some trajectory analysis that you finally matched the observations. The advection between flight legs in opposite direction does not look like two parallel lines. A flight pattern like a sharp triangle would be best to optimize lagrangian observations. In general, some more quantification demonstrating the success of the measurement strategy would help.

Table 4: I'm wondering, why there are "only" seven flights spread over a period of six month. Nothing is said on the selection of these days. Did you specifically wait for the perfect day with the cloud conditions you waited for? Or had you constraints in the available flight hours? I'm wondering what could have been the maximum number of flight with close collocation to the ground sites and how this relates to the seven ECALOT flights.

L 378: Where is the data archived? Does not harm to mention here the data base where users can access the data.

Table 6: I suggest to add the EarthCARE CPR here for a direct comparison. What are the sensitivities of the radars? Could this be quantified in a dbZ threshold?

Section 5.2.1.2: A direct comparison of the specs of AECL and ATLID would be nice.

L 434-435: What does "large discrepancies" mean? Can you quantify! Especially because the Nevzorov probe has a similar high uncertainty and might not be the truth. Also a quantification of how many data were flagged would be interesting to read.

L 446: What do you mean with "geophysical filtering" here? Something like time difference between EarthCARE and aircraft? I don't understand.

Section 5.2.3: Any reference for the FIR-2. Also no uncertainty is provided here.

L 465: Add "water vapor lidar" to DA10. I already forgot what DA10 was.

L 466: Add "microwave radiometer" to MP3000. I don't remember Table 2 here.

L 475: Any uncertainty estimate or reference for the net radiometer? Without uncertainty you can not compare and validate BBR products.

L 477: Why upward and downward irradiance are treated differently? This would bias the net irradiance.

L 484-490: No need for these lines. Of course you have to do this work when converting into GEOMS format.

L 516: R^2 has nothing to do with an agreement! It only quantifies the linearity of the relation between both quantities in case of the linear regression used here. Look at the coefficients of the linear regression (slope and bias) to judge, if both data sets agree.

I also do not see a regression line in Fig. 4c. Are there any biases? What is the root mean square error, etc. There are many more solid statistical tests/parameters to compare such data.

Figure 4: You should also discuss that there is a tendency, that CPR shows low Ze values down to -30 dBZ while the Radar on NAW does hardly reach values below -25 dBZ. Such differences help you to learn something from the comparison.

L526: What is the "the enclosed percentage frequency"?

Section 6.1.2: This section is not even an example. It basically all repeats instrumentation and intended validation what was already introduced in sections 4 and 5. I do not see an additional value here and recommend to remove this section.

Section 6.2: Section title is misleading. This section is mostly on radar doppler velocity. The retrieval of ice cloud properties was not touched. First two sentences (L 538-539) are good examples of general statements that duplicate what already has been discussed earlier. Give no new or substantial information here.

L 543: Launch of EarthCARE?

L 568: For the comparison of the ice cloud retrieval no plot is provided. This is not convincing for readers.

Figure 5: Panel d: Any discussion why EarthCARE shows systematically higher values? Even the sign is often opposite between both instruments updraft vs. downdraft. If you aim to show detailed validation such striking differences need to be discussed.

Section 6.3: Section title suggests "validation". I do not see any validation in this section. Figure

6 only shows airborne observations.

L 583-589: This section only repeats general knowledge. This might be discussed in the instrumentation/validation approach sections. Here it just distracts from the case study.

L 607-619: Again, this part is more a motivation and general knowledge and not the analysis readers may expect in this section.

L 620-633: These lines are related to the flight strategy and should be implemented in the flight strategy section.

L 638: What is a "strong agreement"? You need to quantify and first state how you compare what quantity?

L 652: Again, what is a "good agreement"?

L 653: Why jumping to Fig. 10 and skipping 8 and 9?

L 653: What "simulated" radiative fluxes means here? More details would be required to understand the quality of comparison.

L 671-675: Again, this is more strategy as it should have been already addressed in section 3 or 4.3 before.

Section 6.6.1: This section only provides basic strategy. No examples are shown. Nothing to learn for readers.

Section 6.6.2.: Also for this section, I do not see a value for readers. Instead, I suggest to provide a statistical analysis of the ground-based data with respect to what range of conditions were observed during the period that can be used for EarthCARE validation.

L 743- 751: This is all motivation and should be moved into the introduction.

L 752++: This section addresses the potential for pure model evaluation without EarthCARE. This aspect does not really fit into the objective of the manuscript, what is clearly EarthCARE. Of course all the ECALOT data can also be used for model evaluation and other purposes. I recommend to discuss this shortly in section 7 but not include any example here.