Review of “Persistent EarthCARE underflight studies of the ITCZ and organized convection (PERCUSION): Contribution to EarthCARE Validation” by S. Groß and 23 Coauthors

https://egusphere.copernicus.org/preprints/2026/egusphere-2026-112/

I found this to be a very interesting and worthwhile paper, though it was not quite what I was expected from reading the abstract. The abstract tells me that during the PERCUSION campaign, HALO made “33 passes under the EarthCARE satellite during 30 research flights”. Based on this, I was looking forward to long, comprehensive descriptions of the validation measurements, coupled with multiple analyses of the EarthCARE instruments’ performance relative to these measurements. What I found instead was a well-crafted advertisement for the use of the HALO and PERCUSION data sets in future, more in-depth EarthCARE validation studies. This is not at all a bad thing. Still, I believe the authors should rework their abstract so that it clearly states the scope of the studies presented in the manuscript. The analyses that are described are both well designed and informative, but sufficient only to scratch the surface of a comprehensive validation paper. That said, attempting to validate all four instruments in a single publication is a Herculean task that would no doubt yield a book length manuscript. The multi-sensor validation overview given in the paper is a fine first step. The PERCUSION mission appears to have been meticulously planned to satisfy EarthCARE validation targets. The manuscript is well written, very well referenced, and entirely appropriate for publication in Atmospheric Measurements Techniques.

We thank you very much for your time you dedicated to review this manuscript, for your careful reading and valuable comments that helped us improve the quality of this manuscript. Based on the reviews we revised the manuscript more substantially to better bring out the main points of this study.

Below I’ve attached an annotated version of the authors’ manuscript with (mostly minor) comments sprinkled throughout. I ask that the authors consider all of these in crafting a final draft of their publication.

We have considered all your comments and will give a point-by-point response. Answers are given in italics and bold. Text changes are marked in italics, bold, and blue.

Abstract:

l. 26: sup → sub

We removed sub-tropics.

l. 27: Add a sentence of two describing validation success (or failure) to the abstract. In very general terms, what worked well and what (if anything) did not

We agree that the abstract did not well capture the main objective of the paper or well summarized the findings. We revised the abstract:

‘In May 2024, the Earth Clouds, Aerosols and Radiation Explorer (EarthCARE) satellite was launched. For the first time a satellite combines two active instruments, i.e., the Atmospheric Lidar and the Cloud Profiling Radar, together with two passive instruments, a multi-spectral imager and a broad-band radiometer, on one single spacecraft platform. EarthCARE is thus the most complex satellite mission to date to for collocated aerosol,

cloud, radiation and precipitation measurements. To utilize the data collected by the EarthCARE mission to its full extent and to support and quantify the data quality and measurement uncertainty, careful and holistic validation activities are needed. For this purpose, we set up an airborne instrument payload on the German High Altitude and LOng-range research aircraft (HALO), which is similar to the EarthCARE instrumentation. We used this payload during an extensive measurement campaign in summer and fall 2024 in the tropic and mid- to high-latitudes targeting to validate the EarthCARE measurements and data products early in its commissioning phase. With this manuscript we aim to give a detailed overview of the PERCUSION mission, and to advertise the use of its data in future more detailed validation studies. We give examples of how to use PERCUSION data to approach the validation of all for instruments of EarthCARE as well as of higher level (i.e. multi-sensor) products, and give first confidence in the quality of EarthCARE data.'

Introduction:

l. 48: Langsdale et al., 2025 is not listed in the references section

We added the missing reference.

l. 50f: Consider citing Gimmestadt et al., 2017 here

Thanks for making us aware of this. We added a reference to Gimmestad et al.

l. 60: the LaRC HSRL was also a critically important component of CC-VEX

Thanks for making us aware of this. We changed it as follows:

'deploying an elastic backscatter lidar and a W-band cloud radar on the ER-2 aircraft and NASA's HSRL (Hair et al., 2008) system on the Beechcraft B-200 King Air. '

l. 70: one → ones

Done

l. 77f: one would think the causal connection between the two activities would be reversed; that is, rather than acquiring measurements designed to be "consistent with corresponding [...] simulations", the simulations should be revised to be consistent with the measurements.

Right! We changed the sentence accordingly.

Section 2 – Requirements for validation:

l. 101ff: For each instrument, please include references to the papers that provide descriptions of the engineering specifications and measurement capabilities; e.g., similar to the papers listed in table 3 for the HALO instruments. For CALIOP, a good example of this sort of paper is Hunt et al., 2009 (https://doi.org/10.1175/2009JTECHA1223.1).

We discussed this issue with resoponsible people at ESA. The recommendation we received was to refer to the general overview (Wehr et al.) of the mission and retrieval (Eisinger et al.) and the publication referred to within. We follow their recommendation.

l. 114: in level products → according to data processing level.

We changed that.

l. 123: described by [who? what?] and

Sorry, the reference was missing. We added it.

l. 124: unites → units

Done

l. 124: hyphen is unnecessary

Done

l. 125: the following → Table 2

Done

Table 2 (cloud layer detection): should cloud phase be explicitly included in this list?

We agree that the cloud phase was missing. We added this to the macrophysical properties.

Table 2 (AOT – aerosol types): wouldn't ATLID measurements also be helpful in distinguishing between different aerosol types?

Right! However, we expect, the main problem here is the sensitivity of the passive sensor to the underlying surface. Thus, we want to keep this point separately.

l. 132: are multi-layer aerosol scenes (e.g., dust lofted above marine) part of this category?

To our understanding this includes also multi-layer aerosol scenes. To make that clear we added this sentence to the text.

l. 136: be characterized → characterize

Done

l. 138 (is challenging): perhaps soften this by saying "can be challenging". A-Train observations show that most marine stratocu is readily detected by either CloudSat or CALIOP (though not always at the same time, and, as the authors note, CALIOP frequently fails to detect the true base of optically thick clouds)

Thanks for that advice. We reformulated the text accordingly.

l. 147 (detect): would "identify" be a better word choice here?

Changed it.

l. 153: Combination of radar and lidar → Combined radar and lidar measurements are

Done

Section 3 – The HALO PERCUSION mission

l. 167 can the authors provide a reference describing this activity?

We are not sure what reference would be a valuable contribution here or how to reference our activities. During the PERCUSION campaign (and during following campaigns we performed for validation) our flight planning was closely coordinated with any activities on ESA side regarding the calibration of instruments or maneuvers for orbit adjustments. This close coordination helped to have EarthCARE always fully operational when an underpass was performed.

l. 170 (15 km): good for mid-latitude observations but insufficient for measuring a large fraction of tropical cirrus

That's true; even a flight altitude of 15 km was often insufficient to fully capture cirrus clouds in the tropics. However, convection in the tropics wasn't always fully developed, and we were able to conduct good validation measurements even at these latitudes. However, we are aware that due to this (and also other limitations, e.g. no in-situ measurements) the word 'ideally' is not right here and changed it to 'well'.

l. 174: Doppler should be capitalized

Done

Table 3 (HAMP): please define the acronym

It is HALO Microwave Package

Table 3 (MIRA): please define the acronym

That's the manufacturer's model name. Unfortunately, we can't provide a definition.

l. 183: is it safe to assume that the EarthCARE operational algorithms were also adapted for use with the HALO measurements?

This is true for some but not all of the operational algorithms, e.g. we applied the CAPTIVATE (radar-lidar) retrieval also on HALO radar and lidar data.

l. 186: which one? the reference section lists two Gutleben et al., 2019 papers

Thanks for making us aware. We clarified that throughout the text.

. 195: Next → In addition

The paragraph was rephrased.

Table 4 (water path): ice water path AND liquid water path, or ice water path only?

Our main focus is on the ice water path. We changed that accordingly.

Measurement strategy: holy moly! what a barrage of acronyms!!

Sorry for that ;-)

l. 224: During this part → This

Done

l. 225: it was aimed → was organized

Done

Figure 1: the Sal and Barbados plots are difficult to interpret, as it's not immediately clear what the circles represent. why are there no circles in the Oberpfaffenhofen plot? if possible, add legends that assign flight dates to each color used

Thank you for this comment. We thought of adding the flight dates for each flight / color. However, the plot got too overloaded. We added a description of what purposes the circles were for in the figure caption and in the text.

l. 238f: please include this information in the caption of figure 1

Done

l. 246: catched → flew directly beneath

Done

l. 246: northern → the northern

We changed that to 'in the northbound and southbound direction'.

l. 247: meeting point → closest coincidence

Done

l. 256: cross-track distance, yes?

Yes, we added this.

l. 257: generally, yes. but probably not for small scale boundary layer cumulus.

Yes, this is right, but even for small scale structures we found from previous studies that the aircraft and satellite measurements can be compared to give a meaningful evaluation of height distribution and horizontal extend. We added the following sentence:

'However, care is to be taken when comparing small scale structures, e.g. boundary layer clouds, in those cases.'

l. 264: meeting → coincident

Done

Table 5: reminder on target scenes

We don't know, what is meant here. The targeted scenes are given in the table.

Table 5: on the orbit numbers, what do the C, D, and E postfixes indicate?

C, D, E, ... give information on the latitude of the orbit. The orbits are thus subdivided according to their position.

Section 4 – PERCISON's potential for EarthCARE Validation

l. 271: overview, → overview

Done

l. 271ff: I found this paragraph very useful in bounding my expectations of the material that follows. it would be a big help to other readers if a synopsis of this statement of intent was clearly given in the abstract. (if it's meant to be there, I missed it.)

We have revised the abstract to be more precise.

l. 283: here and elsewhere, please also convert along-track temporals resolution to approximate spatial resolutions

Done

l. 288: sentence fragment

We rephrased this section and integrated the fragment into a longer sentence.

l. 292: using the coordinates in table 5 and looking at google earth tells me that this data segment was acquired over ocean. including an orbit track map would be helpful. explaining that the "signatures of low clouds" are shown in black in both images would also make the info content of these images more readily accessible.

We followed this advice and mention that the measurements were performed over the ocean and that clouds are indicated by black signatures.

Figure 2: consider adding wavelength info to the legends

Due to the space limits in Figure 2 we did not add the wavelength information there, but it is noted in the text and on the 2D Plots, as well as in the legend of Figure 3.

Figure 2: the different wavelengths, the cirrus attenuation, and the different measurement time scales make it difficult to directly compare the backscatter magnitudes. (at first glance ATLID appears to underestimate WALES.) but isn't the depolarization of dust largely independent of wavelength at 355 and 532? so perhaps an additional comparison of the depolarization ratios in the dust layer would yield some useful insights?

As this reviewer states further below, we are providing a comparison of the depolarization ratios. In this plot we focus on the comparison of the Level 1 data. We included sub-headlines to better highlight if we are comparing L1 or L2 data.

Figure 2: why was this comparison not centered around the coincident location shown by the dotted line?

Around the point of best match, the aerosol structure is spatially very inhomogeneous and thus more susceptible to differences on small scales. On the other hand, the signal of ATLID had to be averaged over longer distances to let systematic differences stick out form the noise. Thus, a region as close as possible to the best match time was chosen for which the aerosol layer is as homogeneous as possible. The plot region was extended to better show the inhomogeneous parts on this side.

l. 305: what is "A-NOM"?

ATL_NOM_1B is the product name in the ESA database. We changed A-NOM to the correct name and give the explanation in the text.

l. 312: evaluating of the → evaluating the

Done

l. 315: can the authors comment on any specific metrics that might be applied in future, more quantitative analyses?

This manuscript does not intend to give a full validation. However, we do a careful validation that will be presented in a following manuscript. In our analysis we are doing a case-by-case validation as well as a statistical validation; taking into account the differences in the mean profile as well as of the mean properties sorted by scene, height and feature.

Figure 3: interesting. Floutsi et al., 2023 (https://doi.org/10.5194/amt-16-2353-2023) suggest that lidar ratios for Saharan dust are ~53 ± 8 sr at both 355 nm and 532 nm.

There are several studies addressing the lidar ratio of Saharan dust at different wavelengths form this group as well as of Floutsi et al.. It was found, that the lidar ratio for Saharan dust depends on the dust source region and on dust transport pathway/time. Furthermore, it has been shown from direct comparisons that the lidar ratio derived from HSRL instruments is sometimes found to be lower than the values derived from Raman lidar systems (e.g. Esselborn et al., 2008, Tesche et al., 2009) due to the different averaging needed in for the Raman signals. Since it is not the objective of this manuscript to investigate the lidar ratio of Saharan dust but to evaluate the quality of the derived values form EarthCARE, we decided to keep this paragraph as is. Deploying the same high-quality method (HSRL) we prevent of adding systematical biases to this comparison.

Figure 3 (Lidar ratio): add units

Done

Figure 3: a-ha! here is the depolarization comparison I asked for earlier. how much vertical smoothing is done to generate these plots?

For the WALES data no vertical smoothing is applied, only horizontal over the given average interval. For ATLID the retrieval algorithm decides on its own how to locally smooth the data. The size of these regions can be seen as the blocks of constant color in the 2D plots for the depolarization.

l. 334f: please be more specific; provide more detail on "issues in the layering of the A-EBD properties"

We have rephrased this sentence to be more specific:

'Early comparisons between HALO measurements and ATLID L2 optical properties have identified an error in the layer specification for EarthCARE A-EBD algorithms, resulting in overly coarse layering. This problem could be before data have been made publicly available and are no issue anymore in the current baseline.'

l. 336: please provide examples of these problems and comment on their severity

We have removed this sentence since the information is given now already in the information given. See prior comment.

l. 349: : → .

Done

l. 350: ~16:19?

Done

l. 353: indicative of → indicates

Done

l. 356: these algorithms are used only on the WALES data. yes? aren't the ATLID layer boundaries determined using the algorithm described by van Zadelhoff et al.., 2023?

We agree that the previous wording was not sufficiently clear. For the WALES backscatter data shown in Fig. 2a, a height-dependent threshold-based algorithm was developed to distinguish explicitly between cloud and no-cloud conditions. This algorithm is described in detail in Krüger et al. (2026, in preparation) and is applied exclusively to the WALES observations. Regarding the A-TC product, Irbah et al. (2023) describe that the classification of atmospheric targets is based on height-dependent thresholds applied to ATLID backscatter and depolarization ratio, together with information on the tropopause height.

To clarify this distinction, the corresponding paragraph in the manuscript (ll.356-362) has been revised as follows:

"Threshold-based detection algorithms are commonly applied to determine atmospheric targets from lidar observations such as clouds and aerosols (Groß et al., 2013; Marinou et al., 2019). In this example, however, the focus is exclusively on the validation of EarthCARE cloud products. In particular, we focus on cloud macro-physical properties from the target classification product (A-TC), providing information on the vertical distribution of clouds and the cloud top height product (A-CTH). The A-TC classification itself applies height-dependent thresholds to ATLID backscatter and depolarization together with information on the tropopause height to distinguish between atmospheric target (Irbah et al., 2023). For the validation, cloud pixels and the altitude of the uppermost cloud layer are independently determined from the WALES backscatter ratio (BSR) data using a height-dependent threshold-based algorithm similar to approaches used in previous studies (e.g., Groß et al.,

2014; Urbanek et al., 2017; Gutleben et al., 2019; Dekoutsidis et al., 2024). The specific implementation of this algorithm is described in detail in Krüger et al. (2026, in preparation) and is applied exclusively to the WALES observations."

We also provide the new references included in the text:

Dekoutsidis, G., Wirth, M., and Groß, S.: The effects of warm-air intrusions in the high Arctic on cirrus clouds, Atmos. Chem. Phys., 24, 5971–5987, https://doi.org/10.5194/acp-24-5971-2024, 2024.

Irbah, A., Delanoë, J., van Zadelhoff, G.-J., Donovan, D. P., Kollias, P., Puigdomènech Treserras, B., Mason, S., Hogan, R. J., and Tatarevic, A.: The classification of atmospheric hydrometeors and aerosols from the EarthCARE radar and lidar: the A-TC, C-TC and AC-TC products, Atmos. Meas. Tech., 16, 2795–2820, https://doi.org/10.5194/amt-16-2795-2023 , 2023.

Krüger, K., Groß, S., and Wirth, M.: Validation of cloud macrophysical properties from the ATLID L2a products A-TC, A-FM, A-CTH using airborne lidar observations during the HALO missions PERCUSION, ASCCI and NAWDIC, (to be submitted to AMT, 2026).

Figure 4 (small values): what are the units of the x-axis?

Addressed in the comment below

Figure 4: please explain the significance of the red and gray lines in the figure caption

We thank the reviewer for the suggestion regarding the x-axis. The unit of the lower x-axes is indeed in degrees latitude. To improve clarity, we have updated the figure as follows:

- ***Adjusted the backscatter colorbar by reducing the number of colors and setting new thresholds, corresponding to Krüger et al. (2026, in prep.).***
- ***Added a legend showing cloud tops from WALES (red) and the A-CTH product (gray). Only WALES cloud tops that could be unambiguously identified—i.e., where the distance between the cloud and the aircraft was sufficiently large—are displayed now.***
- ***Included a proper labelling of the lower x-axes in panel (a) and (b) (°N).***

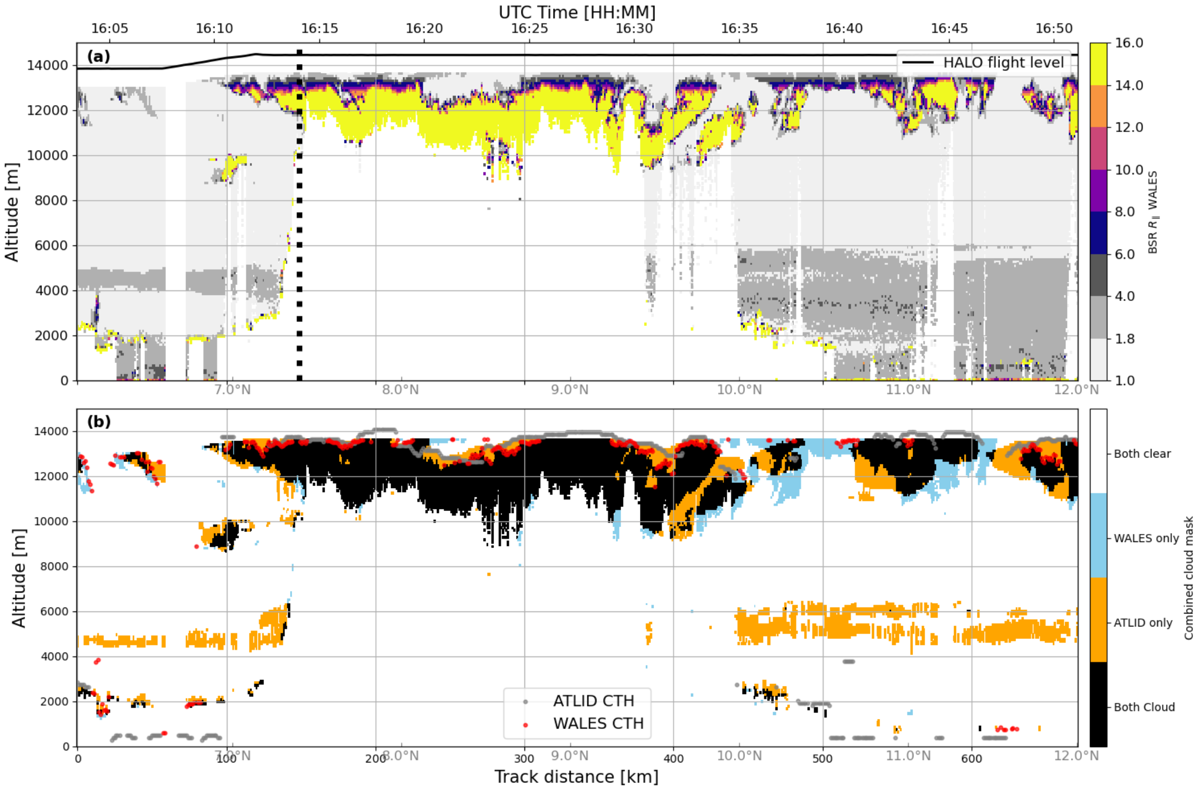
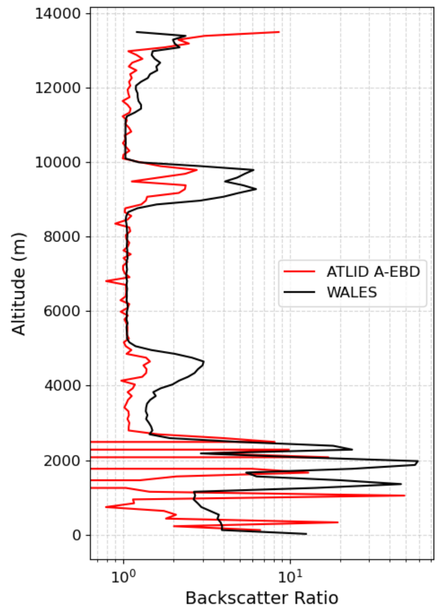


Figure 4: interesting. is the thin area of relatively strong scattering also seen in the ATLID data? or was there a noticeable change in cloud backscatter here between the two sensors?

Thanks for your comment. We tried to improve the figure caption also by including a statement how to interpret and compare WALES and A-CTH cloud top altitudes.

"Figure 4: Distance–altitude cross sections for the HALO flight on 16 August 2024 showing (a) the backscatter ratio measured with the WALES HSRL, along with the flight altitude of the HALO aircraft (black solid line) and the EarthCARE underpass track (thick black dashed line). Panel (b) shows the combined cloud mask derived from the WALES BSR data and the collocated EarthCARE Level-2 target classification product (A-TC). Black shading indicates cloud pixels detected by both WALES and A-TC, while blue and orange colors represent cloud pixels identified exclusively by WALES or A-TC, respectively. White denotes regions with no cloud detection by either product. The red markers indicate the cloud top height derived from the WALES backscatter ratio, while the gray markers represent the cloud top height from the A-CTH product. Note that cloud tops from WALES are displayed only when they could be unambiguously identified (e.g., when the cloud top was sufficiently far from the HALO aircraft), while A-CTH reports a cloud top for these cases."

Figure 4: it would be helpful to have latitude/longitude coordinates on the x-axis rather than along-track distance



We are not entirely certain which specific area the reviewer is referring to. To clarify this point, we calculated the mean backscatter ratio (BSR) profiles for the first 100 km of the flight legs where both instruments observed the same atmospheric features. These segments include (i) ice cloud patches between approximately 9–10 km altitude and (ii) slightly enhanced WALES BSR values around 4-6 km within the aerosol layer.

It should be noted that a direct quantitative comparison of the BSR magnitudes is limited because the instruments operate at different wavelengths (WALES at 532 nm and ATLID at 355 nm). Due to this wavelength dependence of scattering, the BSR peaks observed by ATLID are expected to be smaller than those measured by WALES.

Nevertheless, the vertical positions of the enhanced backscatter are consistent between both instruments. Both instruments show pronounced features at approx. 5 km altitude associated with the aerosol layer, as well as strong signals between about 9–10 km corresponding to the ice cloud.

l. 377: from looking at the images, the tendency that you allude to here is not apparent to me

We realize that these statements could be misunderstood in this context. We report that the ice clouds in the A-TC product are thicker vertically, which generally means that their upper limits are higher. However, the cloud upper limit in A-TC (for example, the uppermost cloud pixel in this product) is not analyzed in this example. Instead, we note that the cloud tops from the independently generated A-CTH are above those from WALES in the high-altitude clouds, which is consistent with the data from A-TC. A detailed and statistical validation of the A-CTH product will follow in Krüger et al., 2026 (in prep.).

l. 378: some comments on WHY this misclassification occurs would be enormously helpful

Thanks for your comment! In general, potential sources of misclassification include issues with averaging and layering procedures (see comment below), limitations in the depolarization channel, and the use of semi-optimal thresholds in the co-/cross-polar and Rayleigh channels by the processor to separate clouds and aerosols.

We agree that, in order to improve the quality of the ATLID level 2 products, it is necessary not only to detect misclassifications but also to identify their underlying causes. However, the aim of this section is to demonstrate the potential of the WALES HSRL data set for validating ATLID Level 2 cloud products. Identifying the exact causes of misclassification is therefore beyond the scope of this paper and a more robust, and statistical validation approach based on more cases, and would benefit of the involvement of the algorithm developers.

l. 383: plus a hefty dose of false positives. so presumably the A-TC product under-represents aerosol presence in this scene.

We agree that the overestimation in ice cloud pixels in this case may cause an underrepresentation of aerosol pixel in this case. We decided to emphasize this point more clearly and added a sentence in the manuscript.

"For this case study, this type of misclassification leads to an overestimation of cloud pixels and presumably a corresponding underestimation of aerosol pixels within this altitude layer."

l. 385: this is the second time I've encountered this phrase. please explain what it means.

By "bad layering," we refer to Section 2.1.1 of Irbah et al., 2023. This section describes that, for generating the target classification, the Feature Mask is used as input. The Feature Mask (A-FM, van Zadelhoff et al., 2023) serves as the starting point for ATLID Level 2 processing. Its purpose is to provide the best possible strategies for optimal averaging (e.g., determining boundaries between clear-sky and optically thin pixels), which are then used to generate further downstream Level 2 products. If transitions between different flags are less sharp, this can lead to suboptimal processing (for example causing optically thin clouds to be misclassified as aerosol). For further details, we refer to Irbah et al., 2023.

We have tried to clarify this sentence somewhat and have decided not to use the expression "bad layering" as follows:

"There is indication that cloud pixels in A-TC, especially ice clouds, appear too large horizontally and vertically, which could be due to suboptimal averaging in ATLID Level 2 processing (see Irbah et al., 2023)."

l. 386: as a prospective data users, I'd very much like to know which product is the more reliable (and why)

We agree that it would be useful for the user to know which product is more reliable. However, this is beyond the scope of the present paper, which only presents the cloud distribution (A-TC) and cloud tops (A-CTH) products for a single case. Drawing conclusions about the reliability of individual products requires a careful statistical validation. Such a validation study is currently in progress (Krüger et al., 2026, in prep.).

l. 388: using the CPR acronym to refer to the radars on both platforms could get a bit confusing for readers.

Right. We now use 'MIRA' or 'radar onboard HALO' if we refer to the airborne radar.

l. 395: here and elsewhere, Doppler should be capitalized

Done

Figure 5: image annotations are blury (perhaps they originated as highly compressed JPEGs?)

We provided an improved version.

Figure 5 caption: Earth CARE → EarthCARE

Done

l. 404: are these anomalies also present in the standard data products or are they largely confined to this example?

These anomalies are a feature that appeared during the commissioning phase. They are not present in the provided data products. We added the following sentence to make that clear:

'This led to the apparent Doppler gradient in clear-sky regions (Figure 5c) biasing low echo cloud regions which disappeared after switch to the operational signal processing unit (SPU-B) on 4th December 2024.'

l. 409: to the uninitiated, it's not at all clear what BA, CA, and CB are supposed to indicate. are they data processing versions? if so, please provide a pointer to a document that explains the nomenclature.

Right, BA, CA, ... indicate the processing versions. Unfortunately, there is no overview of all release notes, however, we added a link to the current release note for ESA products (EarthCARE Data - Earth Online) and JAXA products (EarthCARE - Earth Clouds, Aerosols and Radiation Explorer).

l. 420: Ewaldet → Ewald et

Done

l. 425: how close is close enough in space? according to the text around line 257, the instruments were within 500 m of each other, except for the two first flights, where they were within ~1000 m.

This is a very general question and cannot be fully answered, as it depends on the variability of the scene, the size of the features being examined, and the viewing angle. The point being made here is that, due to the different viewing geometry of the instruments, a direct comparison can only be made with respect to the nadir pixel, regardless of how precisely EarthCARE was underflown.

Figure 7: IMO, all of these images would benefit from using larger fonts for the axis labels, titles, and legends

We enlarged all font sizes and lines.

Figure 7 caption: in b → panel b,

We applied all suggested changes.

l. 433: "tropical North Atlantic" is an oxymoron (the North Atlantic is certainly not in the tropics!)

Thanks, we changed this to ‘northern tropical Atlantic’.

l. 434: radiances

This has been applied.

l. 443: unclear; isn't this whole orbit segment over the ocean? do you specifically mean cloud-free skies over the ocean?

We clarified this and changed the sentence to: Good agreement between the two instruments is particularly also observed for clear-sky pixels over the ocean…

l. 456: have there been any attempts to use MSI data to help characterize ATLID multiple scattering in cirrus, as in Garnier et al., 2015 (https://doi.org/10.5194/amt-8-2759-2015)?

We are not sure, how this question refers to the validation of MSI radiances. The ATLID retrieval and multiple scattering treatment are published by Donovan at al., 2024. We use the HALO measurements during PERCUSION (and following HALO campaigns with the WALES system) to characterize ATLID multiple scattering in cirrus using different multiple-scattering codes and different retrievals for the assumed particle size.

Figure 8: this map makes a very welcome contribution to this figure. maps would be equally helpful for figures 2, 4, 5, and 7.

We agree that including the underpass location on a map would provide valuable information; however, the main objective of this manuscript is to provide all necessary information about the PERCUSION campaign and its use for EarthCARE validation. Therefore, we will refrain from providing a map for each example. We will, however, gladly consider this suggestion for the individual validation papers.

BBR: I don't have much insight to offer on this topic, so I hope some other reviewer will be confident to evaluate the material.

l. 498: perhaps add another appendix that decodes the various data product names (does BMA_FLX_2B have an English translation?) and explains the differences between the data processing versions (.e.g., AB)

Thank you for this comment. We considered following this reviewer's advice but ultimately decided against it, as a truly comprehensive explanation would be too lengthy for this manuscript. The data products are described in detail by Eisinger et al. We refer to this publication and description in the manuscript. We furthermore added a link to the current release notes from ESA.

l. 493: released with → reported in

Done

l. 494: product → product, available

Done

l. 516: hight → height

Done

l. 526: the → an

Done

l. 526: estimate → estimation

Done

l. 543: delete ‘will be’

Done

l. 544: acronym already defined

Right. Changed.

l. 586: sentence fragment

Corrected

l. 660: change to final version

Changed

l. 728 (Groß et al., 2014: missing DOI

Corrected

l. 883: is → are

Corrected

l. 887: seve → serve

Corrected

l. 890: date and time → dates and times

Done

l. 899: Earth CARE → EarthCARE

Done

l. 903: → measurements,

Done

l. 907ff: suggested revision: However, our combined active (lidar and radar) and passive remote sensing payload (Stevens, et al., 2019) bridges both satellite missions with correlative airborne lidar and radar measurements underflying both CALIPSO/CloudSat and EarthCARE.

Done

l. 909: unterflight → underflights

Rephrased

Table B1: are these data also available via a Zenodo link?

We uploaded the basic lidar and radar properties as for the WALES lidar data and the MIRA cloud radar data from A-Train underflights with HALO as two combined datasets on Zenodo:

'The basic lidar and radar products are analyzed and available as two combined datasets for WALES lidar data and MIRA cloud radar data from A-Train underflights with HALO via Zenodo (Ewald and Wirth, 2026; Ewald, 2026).'

Ewald, F.: MIRA cloud radar data from A-Train underflights with HALO (Radar reflectivity) [Data set]. Zenodo, https://doi.org/10.5281/zenodo.19317830, 2026.

Ewald, F., and Wirth, M: WALES lidar data from A-Train underflights with HALO (Backscatter coefficient) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.19318124, 2026.

Summary Comments

The manuscript provides a detailed description of a series of airborne campaigns deploying a suite of instruments assembled to assess and validate all instruments on the EarthCARE satellite. The paper provides a concise overview of the EarthCARE objectives and clearly articulates the validation measurement needs of all four instruments on the EarthCARE satellite over a range of atmospheric conditions highlighting the uniqueness of the instrument suite deployed. The introduction was well written, and provided a clear description, application, and impact of the using airborne instrument suite for validation. In contrast, the abstract should be changed to be clear on the overall objective or purpose of this manuscript and provide a summary statement of the conclusions. For example, the manuscript describes the rationale, comparison approach for each instrument, and shows initial comparisons thus providing a substantial contribution to the evaluation and validation of EarthCARE's measurements. These comparisons start to build confidence in the EarthCARE measurements and data products for use by the science community and are already providing feedback to the algorithm developers to adjust processing methods and identifying additional calibration needs.

The aircraft remote sensing instruments deployed are clearly described in the manuscript with appropriate references used for calibration, demonstration of their heritage, and are of high quality. The measurements provide a capable and unique suite of measurements that was deployed in multiple geographic regions with varying atmospheric conditions of clouds and aerosols and spanned both daytime and nighttime as presented in the examples for each instrument. The paper clearly states, although not in the abstract, that the intent is not to provide an exhaustive validation analsysis but rather illustrate how the data has been and will continue to support assessments of EarthCARE's measurements and retrieval algorithms for key geophysical variables.

There are two appendices that came across as after thoughts in the manuscript. The first appendix provides specific flights coordinated underflights of NASA's PACE satellite during the deployments discussed in the manuscript (see specific comments on the flight track figures below) and is only discussed in the conclusions section at the very end. The second appendix is a summary of the previous coordinated flights with NASA's CALISPO and CloudSat satellites and similarly has only a few sentences at the end of the conclusions. In this case, the authors postulate that airborne flight campaigns can link the two separate datasets without providing specific details on how that might be done which is likely not trivial. It is recommended to remove this from the conclusions and compile all the tables (PERCUSSION, PACE, CALIPSO/CloudSat) of the under-flight information in a single supplement rather than include these as appendices. These tables can then be used as a reference for future more comprehensive validation efforts.

Lastly, these data will only be useful the larger science community if the datasets are available and open access. A location and description of how to download the CALIPSO/CloudSat flights was not clearly listed in the appendix or in the text. In addition, the data should be processed data (i.e. geophysical data products rather than raw signals) if they are to be fully exploited for validation by the larger community. Overall, the presentation quality is generally good. The manuscript has value to the larger community to better under this unique dataset with specific examples presented to address all four EarthCARE instruments. Specific detailed comments below can enhance clarity, particularly in sections discussing lidar measurements.

We thank you very much for your time you dedicated to review this manuscript, for your careful reading and valuable comments that helped us improve the quality of this

manuscript. Based on the reviews we revised the manuscript more substantially to better bring out the main points of this study.

We have considered all your comments and will give a point-by-point response. Answers are given in italics and bold. Text changes are marked in italics, bold, and blue.

Specific Comments/Suggestions:

- Provide clear statement of the overall objective of the manuscript and summarize conclusions in the abstract. It is not clear from the abstract what is the main objective paper and there is no conclusions/im pact statement provided.

We agree that the abstract did not well capture the overall objective of the manuscript. We revised the abstract:

'In May 2024, the Earth Clouds, Aerosols and Radiation Explorer (EarthCARE) satellite was launched. For the first time a satellite combines two active instruments, i.e., the Atmospheric Lidar and the Cloud Profiling Radar, together with two passive instruments, a multi-spectral imager and a broad-band radiometer, on one single spacecraft platform. EarthCARE is thus the most complex satellite mission to date to for collocated aerosol, cloud, radiation and precipitation measurements. To utilize the data collected by the EarthCARE mission to its full extent and to support and quantify the data quality and measurement uncertainty, careful and holistic validation activities are needed. For this purpose, we set up an airborne instrument payload on the German High Altitude and LOng-range research aircraft (HALO), which is similar to the EarthCARE instrumentation. We used this payload during an extensive measurement campaign in summer and fall 2024 in the tropic and mid- to high-latitudes targeting to validate the EarthCARE measurements and data products early in its commissioning phase. With this manuscript we aim to give a detailed overview of the PERCUSION mission, and to advertise the use of its data in future more detailed validation studies. We give examples of how to use PERCUSION data to approach the validation of all for instruments of EarthCARE as well as of higher level (i.e. multi-sensor) products, and give first confidence in the quality of EarthCARE data.'

- Recommend moving Table 5 to a supplemental section along with other flight information tables for the PACE and CALIPSO/CloudSat.

We agree that the information in the appendix appears somewhat unintroduced and unmotivated (with only a brief mention in the Conclusions). We include now the information of underflights under PACE mission during the second part of the PERCUSION campaign, and the information to CALIPSO/ClouSat already in the PERCUSION section to better introduce this addional objective.

'In addition to our primary goal of validating EarthCARE, our efforts also address the validation of measurements and algorithms from NASA's PACE (Plankton, Aerosol, Cloud, Ocean Ecosystem) mission. To this end, we conducted four dedicated flights in the PACE swath during the campaign phase over the tropical western Atlantic. Furthermore, in preparation for EarthCARE, we performed a series of underflights with the same payload under NASA's mission constellations CALIPSO and Cloudsat. Together with the measurements for EarthCARE validation, these underflights offer the opportunity to bridge some of the gaps arising from the two missions, such as those concerning wavelength

dependencies and sensitivities. Both topics will be addressed in the appendices of this manuscript.'

We do not want to put the information about the conducted underflights into a supplement, since we believe the detailed information about the underflights is an important part of this manuscript.

Figure 2 and discussion therein:

a) Care should be taken with the terms backscatter, backscatter ratio, and backscatter coefficient. For example, Rayleigh backscatter profile is used but not actually defined. Care should also be taken when comparing backscatter coefficient attenuated products with unattenuated products. For example, all of the Reyleight products are attenuated as shown.

   We agree that care should be taken here. Thus, we went through the text carefully. We introduced the term backscatter ratio where used the for the first time to clearly distinguish between backscatter coefficient and backscatter ratio. We also changed the phrasing from e.g. Rayleigh backscatter profile to Rayleigh backscatter coefficient to have no inconstancies between the figure and the description. However, we do not quantitatively compare the attenuated products to unattenuated products. The time-height cross-sections from WALES and ATLID are shown to give some context of the scene and confirm its stability.

b) Labels in upper panel have label of unpolarized – caption states this to be co-polarized and should be changed.

   We had a critical look again at figure label and figure caption and confirm, that the caption and the labels of the time-height cross-section is correct as given. The time-height cross-sections show the unpolarized (attenuated) backscatter coefficient for both instruments to give the context of the observed scene. In the profile comparisons we look at Mie, Rayleigh and cross-polar backscatter coefficient. Please be aware, that this is not the same variables as shown in the time-height cross-sections.

c) Why does the comparison not include the coincident location or be centered at the coincident location? Please explain/justify. Text refers to atmospheric conditions as being stable.

   Just around the overpass point the aerosol is very structured, which enhances the probability for representativeness errors. We extend the text to explain this better.

d) Why was 120 km used for the comparison? Please provide a rationale.

   The single profile data from ATLID is quite noisy, so a large averaging interval was chosen to reduce the statistical noise and make systematic effects stick out.

e) No error bars provided for this comparison. Provide careful analysis to include error estimates for both datasets.

The error bars for the statistical noise are so small that they were hardly visible. We added them now and you can see them if you zoom in.

f) No statistical comparison is provided between the two different profiles (e.g. mean differences etc.).

We agree, that doing a statistical comparison is very valuable. However, in this manuscript we only show one single example to advertise the use of PERCUSION for EarthCARE validation and give the main information. The statistical basis from one example is too low. We are preparing a manuscript targeting ATLID L1/L2 validation with the WALES measurements during PERCUSION where we include all 33 underpasses. We will add a statistical evaluation in this more detailed study.

g) The L1 ATLID data products are attenuated backscatter for the three components of scattering rather than unattenuated backscatter which is shown and therefore I suggest not stating this as a 'direct' L1 comparison. Showing and stating that a direct comparison of the attenuated ATLID product from the WALES data (unless a wavelength dependent attenuation correction is applied) is confusing. Although a cross reference (i.e. normalization to the ATLID data) is done for the WALES Rayleigh attenuated backscatter it is not clearly stated why this is required for comparison for the average reader. There are terms like 'extinction corrected' and 'aligned' which are not defined and discussed to the why they are used.

As already mentioned for comment b), the time-heigh cross-sections are intended to give the scene context (geometry, homogeneity). The actual comparison is done on the co-polar, Rayleigh and cross-polar backscatter coefficient. For these quantities, of course, the different attenuation by molecular scattering for the different wavelengths has to be corrected.
This is done using the air density profiles given within the products. 'Direct' comparison here means that averaged profiles close to the raw L1 data are compared. As we see this might lead mis-interpretation we removed the word 'direct'. Furthermore, we already mention in the text, that we have to correct for different attenuation due to the different wavelengths; thus the data has to be extinction corrected for the comparison. ATLID data (co-polarized backscatter coefficient) is not given extinction corrected (ATLID raw) we use the information in the Rayleigh signal to correct the attenuation to make the profiles comparable to WALES co-polar backscatter coefficient, which is fully corrected for extinction. I that might not have been clear, we changed the text as following:

‚The backscatter coefficient for co-polar polarization for WALES and for ATLID is shown without extinction correction as contained in the L1 product and with correction for aerosol and molecular extinction by dividing the co-polar signal by the Rayleigh signal (HSRL method, purple line). The WALES profile shown is fully corrected for extinction.'

Since it was also not clear why we had to align the profiles for the Rayleigh backscatter coefficient, we added the following information:

'For evaluating the ATLID Rayleigh backscatter coefficient, we take into account the expected pure Rayleigh backscatter signal calculated from an atmospheric density profile contained within the L1 data and this pure Rayleigh signal at 355 nm with additional aerosol extinction derived from the WALES measurements.'

h) Suggest rewriting the discussion and figure caption being more precise on the specific terms and provide clear and concise description of the processing done for the comparisons. Explain why you are comparing unattenuated backscatter coefficients for the Mie components and why you are comparing attenuated backscatter coefficients for the Rayleigh scattering along when a scaling or normalization for above aircraft attenuation is being applied or not being applied. It might also help to define the ATLID L1 products as attenuated backscatter from top of the atmosphere. It looks like these comparisons are using the ratios o L1 data profiles. Again please clearly explain in the text.

We have carefully looked at the description of the processing and added some additional explanations.

i) Is the same molecular density profile used for the air density in WALES and ATLID data?

The ALTID and the WALES data use the air density profiles provided within their products which both are derived from ECMWF IFS analyses. The difference which comes from different interpolation schemes is below 0.5% for the case shown here and does not lead to deviations visible in the plotted data. This will be addressed in the detailed study and publication on L1/L2 validation.

Figure 3 and discussion therein:

a) Why is a 100km averaged used? Please provide rationale.

Please see answers to Figure 2.

b) What is the shading in the line plots (particle extinction and particle linear depolarization)

This is the error estimate given within the data products. We added text to the caption to explain this.

c) Provide error bars for data

Please see answer above.

d) No statistical comparison is made between the measurements is presented.

Please see the answer for point f) in the comments to figure 2.

e) It looks like there is a ~30% difference of the lidar ratio (mostly due to differences in backscatter it appears) and no comment is provided on these differences. In particular, it is stated that the lidar ratio and extinction are expected to be the same. Is this due to measurement error or differences in the wavelengths for this case? Please comment.

We agree that we have been short in the description of this comparison, especially as large differences occur, e.g. for the lidar ratio. We thus extended this paragraph:

'Our comparisons confirm the good quality of the A-PRO products in the current baseline version in general. But differences between the WALES and ATLID data of the different products are visible. While the particle extinction coefficient shows a good agreement within the uncertainty range, significant differences occur for the values of the lidar ratio in the height range between about 1-4 km. The lidar ratio retrieved from ATLID data is about 60 sr while the lidar ratio derived from WALES measurements is about 45 sr. This difference might result from the differences in the derived backscatter coefficient. The discrepancies might result from a wrong target classification and thus a wrong a priori information in the optimal estimate retrieval to derive A-EBD products, or a too strong weight on the a priori information. Investigations about this issue are ongoing, and the problem is hopefully solved in the new baseline version. In addition, for the example shown, significant noise on the depolarization values is observed. Furthermore, especially the depolarization ratio shows a strange behavior at the upper edge of the aerosol layer leading to high, unphysical values.'

f) It looks like ATLID does not have backscatter coefficient measurements above 5 km. Is this correct?

The data within the ALTLID L2 files are flagged as bad (too low SNR) and only fill_values are provided, so yes there is no ATLID data above.

g) The only plot on the bottom panel that has data above ~5km is the first one and there is no ATLID data for comparison. Suggest expanding the scale to show the range of measurements better.

The plots were chosen to cover the range where WALES data is present. That there is no WALES data above 5 km for the more SNR sensitive products at least shows that the aerosol load in this altitude region is so low that the lack of ATLID data can be attributed to this fact.

Figure 4 and discussion therein:

a) Overall the discussion can be simplified which should help highlight the key points when describing layers. For example, "vertically extended layer" is not needed and "indicative of a marked" are extra words that make this hard to follow. Suggest spending time make this discussion starting at line 340 more concise for the readers.

Thank you for this comment. Following the reviewer's suggestion, we revised the paragraph starting at line 340 to simplify the wording and make the description of the observed layers more concise and easier to follow. The paragraph now reads as follows:

'Besides the L2 optical properties we use the WALES measurements for a comparison with the ATLID Target Classification (EarthCARE ATLID TC Level 2A) and the ATLID cloud top height (EarthCARE ATLID CTH Level 2A) (Wandinger et al., 2023) (baseline BA). An example

comparison is shown for the HALO underflight on the 16 August 2024 southwest of Cabo Verde (orbit 01240E). This flight was selected as it captures a wide range of cloud and aerosol conditions typical of the tropical atmosphere. The flight objective aimed at a south-north transect across the ITCZ, including a flight leg of about 700 km (50 minutes) flown along the predicted EarthCARE ground track, extending from the ITCZ centre toward its northern edge. Figure 4a shows a time-altitude cross section of the backscatter ratio (BSR) measured with the WALES instrument with the EarthCARE underpass at 16:14 UTC. The BSR measurements reveals several cloud and aerosol features that are captured in this flight section. At the highest altitudes (10-13 km) frequently large BSR values indicate a widespread field of ice clouds. This ice cloud field has the largest vertical extent near the underpass and is more patchy and thinner toward the northern part of the transect. At lower altitudes (1–3 km), smaller-scale regions of large BSR values are detected which are associated with stratocumulus and cumulus clouds. Additionally in the lowermost 6 km, an extended layer characterized by moderate BSR values (BSR between 2-6) is observed which indicates an aerosol layer. Notably, this aerosol layer appears thicker and more continuous in the northern segment (distance > 400 km) of the flight leg. In the southern part, it appears as two separated layers: an elevated layer around 5 km altitude and another layer within the lowest 2 km of the atmosphere'

b) So all the layers are identified by thresholds it seems from the text which is a common approach for lidar layer detection algorithms. Are these thresholds the same for both ATLID and WALES? If not, one should expect diLerences. Please provide a discussion on the thresholds used for this comparison.
c) Similarly, the averaging scales will also impact these comparisons.

Addressing b) and c)

Thank you for this comment. We would like to emphasize here that the thresholds, methodologies, and averaging strategies used for the generation of the A-TC product and our WALES dataset differ fundamentally. However, our approach is not to mimic the ATLID measurements or the processor algorithms but to provide a fully independent validation. We generally want to highlight in this paper the suitability of the WALES data set for ATLID validation.

In our study, we intentionally follow a fully independent approach by using independent datasets and developing an independent algorithm for cloud detection. Several previous studies have demonstrated that the backscatter ratio at 532 nm can reliably be used to identify and separate clouds and aerosols across different altitude/latitude ranges (Groß et al., 2014; Urbanek et al., 2017; Gutleben et al., 2019; Dekoutsidis et al., 2024).

In addition, the comprehensive A-TC algorithm (see Irbah et al., 2023) relies on individual thresholds for the co- and cross-polarized Mie channels as well as the Rayleigh channel. The algorithm also uses the tropopause height as input and applies different averaging strategies guided by the A-FM product (see Donavan et al., 2024; van Zadelhoff et al., 2023).

Concerning your comment: If, however, the same thresholds were applied, for example to the backscatter ratio, a larger number of cloud pixels would be detected in the WALES dataset. This is due to the wavelength dependence of the BSR (ATLID: 355 nm vs. WALES: 532 nm), which generally results in lower values for ATLID compared to WALES. This effect would lead to a systematic underestimation of cloud pixels by the A-TC products, but the opposite seems to be the case, so the chosen backscatter threshold does not seem to be the critical factor here. We see (Figure 4b) that the A-TC product increases the cloud area

especially of ice clouds, e.g., by filling in gaps (which are detected in WALES). However, these cloud gaps are visible in the A-FM product as well as in Level 1 products, which suggests that the increase in cloud size (as well as a systematic overestimation of cloud pixels) is caused by the processing rather than by an incorrect choice of thresholds.

Finally, we note that a dedicated study is currently in preparation (Krüger et al. 2026, in prep.), which will investigate both the sensitivity of cloud pixel detection to the applied thresholds and the impact of different spatial resolutions (i.e., averaging scales of the A-TC product).

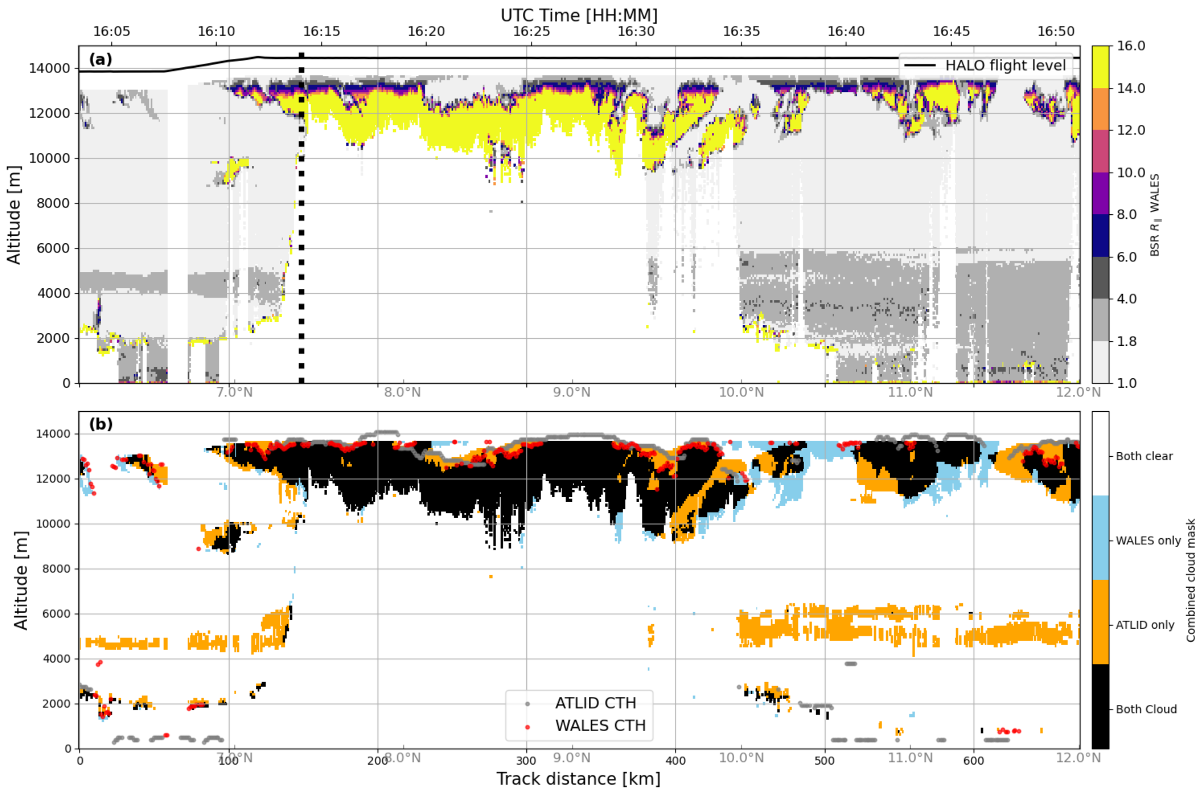
d) I did not see a reference to the CTH lines for the colors (i.e. which one is for ATLID and WALES)
e) For WALES, does the cloud top height need to be below where the data starts? It seems like it is capped at the top of the highest altitude of the dataset (if WALES is the red color)

Comment regarding d) and e):

Thanks for the hints. We generally updated the figure and its caption to improve its clarity (also corresponding to requests from reviewer 1). The updates are:

- ***Adjusted the backscatter color bar by reducing the number of colors and setting new thresholds, corresponding to Krüger et al. (2026, in prep.).***
- ***Added a legend showing cloud tops from WALES (red) and the A-CTH product (gray).***
- ***Included a proper labelling of the lower x-axes in panel (a) and (b) (°N).***

We hope these changes also help the reader to better distinguish between cloud tops from WALES and A-CTH. To remove the "capped" cloud tops from WALES, we now exclusively display WALES cloud tops that could be unambiguously identified—i.e., where the distance between the cloud and the aircraft was sufficiently large—are displayed now.



"Figure 4: Distance–altitude cross sections for the HALO flight on 16 August 2024 showing (a) the backscatter ratio measured with the WALES HSRL, along with the flight altitude of the HALO aircraft (black solid line) and the EarthCARE underpass track (thick black dashed line). Panel (b) shows the combined cloud mask derived from the WALES BSR data and the collocated EarthCARE Level-2 target classification product (A-TC). Black shading indicates cloud pixels detected by both WALES and A-TC, while blue and orange colors represent cloud pixels identified exclusively by WALES or A-TC, respectively. White denotes regions with no cloud detection by either product. The red markers indicate the cloud top height derived from the WALES backscatter ratio, while the gray markers represent the cloud top height from the A-CTH product. Note that cloud tops from WALES are displayed only when they could be unambiguously identified (e.g., when the cloud top was sufficiently far from the HALO aircraft), while A-CTH reports a cloud top for these cases."

Please note the references that were mentioned above:

Dekoutsidis, G., Wirth, M., and Groß, S.: The effects of warm-air intrusions in the high Arctic on cirrus clouds, Atmos. Chem. Phys., 24, 5971–5987, https://doi.org/10.5194/acp-24-5971-2024, 2024.

Irbah, A., Delanoë, J., van Zadelhoff, G.-J., Donovan, D. P., Kollias, P., Puigdomènech Treserras, B., Mason, S., Hogan, R. J., and Tatarevic, A.: The classification of atmospheric hydrometeors and aerosols from the EarthCARE radar and lidar: the A-TC, C-TC and AC-TC products, Atmos. Meas. Tech., 16, 2795–2820, https://doi.org/10.5194/amt-16-2795-2023 , 2023.

Krüger, K., Groß, S., and Wirth, M.: Validation of cloud macrophysical properties from the ATLID L2a products A-TC, A-FM, A-CTH using airborne lidar observations during the HALO missions PERCUSION, ASCCI and NAWDIC, (to be submitted to AMT, 2026).

Figure 7 and discussion therein

a) The lines on the top panels are difficult to see. In fact, the coincident point was not noticed until referenced in the text and is difficult to see (especially when printed and not on a monitor). Suggest making lines more distinct.
b) The colors in lower panels are also difficult to see. The labels are two small to read when printed.
c) Suggest adding quantitative analysis for these comparisons (mean bias, slope biases, etc) with a simple correlation plot?

We used larger font and label sizes and also thicker lines. We did not add another plot in the manuscript, but added quantitative numbers (mean bias and correlation coefficient) in the discussion of the two lower panels:

'For the 670nm channel, we find a mean bias and standard deviation of (-1.28 +/- 6.93)%, and hence slightly smaller radiances of MSI compared to specMACS, and a correlation coefficient of 0.93 for the entire segment. For the 865nm band, we get (1.95 +/- 7.64)% for the mean bias and 0.93 for the correlation coefficient. For the SWIR channels, we find similar biases of (-2.20 +/- 8.51)% for 1650nm and (1.94 +/- 8.63)% for the 2210nm channel, however, they show slightly larger standard deviations. For the correlation coefficients we derive 0.92 for both channels underscoring the good agreement between the two instruments taking into account that we include time differences of around 10 minutes between the specMACS and MSI measurements.'

Figure 8 and discussion therein

a) Using white points on the bottom is difficult to see the points, which are of most interesting being more close in distance (i.e. suggest you highlight the better points to compare).
b) Figure caption, “close to the underpass” is redundant and can be removed.
c) Why are the histograms shown for the lower panel plot? No discussion is provided on the value to include them.
d) Suggest adding addition statistics such as mean bias for this segment of data.

Comment regarding a), c): We changed the color bar indicating the distance from the underpass accordingly and removed the histograms in the lower panel plot.

Comment on b) – We removed “close to the underpass”.

Comment on d) – We calculated the mean bias of the swath averages of MSI and VELOX, which is about 0.15 K. We added this information in the text:

“A mean bias for the swath averages of MSI and VELOX brightness temperatures was found to be about 0.15 K.”

Editorial Comments/Suggestions:

line 53 – change “proofed” to “proved”

Done

line 194 – prefer to use “The specsMACS…” at the beginning of a sentence rather than starting with a lower case acronym (see other places in manuscript).

Done

line 234 – remove “once”

Done

line 245 – change “catched” to “flew under”

Done

line 245 – change “…twice, one in northern direction and once in the southern direction of HALO.” to “in the northbound and southbound directions.”

Done

Line 248 – remove “already”

Done

Line 291-292 – rewrite this sentence as it is not well constructed and is confusing.

We reformulated this sentence and the sentences be for and after as follows:

‘Besides this aerosol layer, signatures of low clouds from about 0.5 km to about 2 km altitude are visible in the measurements of both systems. In addition, enhanced backscattering at around 15 km altitude can be seen in the ATLID measurements, this layer

is missed by the WALES measurements due to the lower flight altitude of HALO of about 11 km.'

Line 324 – "After confirming good performance…" add "…for this day".

Done

Line 347 – Backscatter Ratio is now used and I do not think it has been defined.

We added:

'which is the ratio of the total (aerosol + molecule) to the molecular backscatter signal,'

Line 377 – not sure what is meant by "tendentially vertically thicker"

We agree that this was not well-phrased. We changed it to

'This might lead to vertically thicker cirrus clouds in the A-TC, corresponding to the elevated cloud tops (red line) of the A-CTH product.'

Line 385 – what is "bad layering"?

By "bad layering," we refer to Section 2.1.1 of Irbah et al., 2023. This section describes that, for generating the target classification, the Feature Mask is used as input. The Feature Mask (A-FM, van Zadelhoff et al., 2023) serves as the starting point for ATLID Level 2 processing. Its purpose is to provide the best possible strategies for optimal averaging (e.g., determining boundaries between clear-sky and optically thin pixels), which are then used to generate further downstream Level 2 products. If transitions between different flags are less sharp, this can lead to suboptimal processing (for example causing optically thin clouds to be misclassified as aerosol). For further details, we refer to Irbah et al., 2023.

We have tried to clarify this sentence somewhat and have decided not to use the expression "bad layering" as follows:

'There is indication that cloud pixels in A-TC, especially ice clouds, appear too large horizontally and vertically, which could be due to suboptimal averaging in ATLID Level 2 processing (see Irbah et al., 2023).'

Line 401 – suggest to reference the figure in order (a) then (b).

Done

Line 404 – define IQ oLset

Since we want to provide an overview to a broader audience, we removed the technical term "IQ offset" and replaced it with the more descriptive explanation

'As this measurement was acquired during the commissioning phase, the redundant signal processing unit (SPU-A) of CPR was used which exhibited a drift in the Doppler velocity background signal (see clear-sky regions in Figure 5c), biasing low echo cloud regions which disappeared after switch to the operational signal processing unit (SPU-B) on 4 December 2024.'

Line 432 – should this be 84 m or 84 km?

HALO had a match distance of 84 m from the EarthCARE track during this underflight.

Line 543 – change ‘will be’ to ‘is’

Done

Line 555 -suggest using ‘finer’ resolution rather than ‘larger’ resolution

Done

Line 558 – remove ‘also’

Done

Line 575 – Suggest rewording, “comparison…can confirm the good quality if ATLID

L1b”. This sentence combines that the airborne has the potential to confirm the quality but as stated it infers that this has been done already. Suggest better wording of this statement to reflect that initial measurements have shown qualitative agreement and has been used to identify know issues in the early processing of the ATLID data.

We have generally reworked the ‘Summary and Conclusion’ including this paragraph:

‘A definitive statement about the validation is difficult because the processing of the EarthCARE and HALO data is updated as we gain a better understanding of the measurements. Hence validation is a moving target. For instance, at the very beginning of the EarthCARE mission, PERCUSION measurements helped identify cross-talk effects, background suppression or calibration uncertainties that improved the first release of EarthCARE data products. of ATLID L1b data. It also helped to identify artefacts resulting from bugs in the layer assignment in the algorithm for the ATLID L2a optical products. Airborne radar data from PECUSION measurements helped to identify an offset of the CPR compared to MIRA measurements, and motivated a new calibration of the radar, which led to a much-improved data quality. With the help of the airborne data these problems could be solved quickly and the products publicly available have been updated.’

Line 576 – suggest changing …”are not affected from this issue anymore” to “… have been updated”.

Done