https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

Airborne Validation of Cloud Spatial Properties from the ATLID Instrument on EarthCARE

Natalia Roldán-Henao[1], John E. Yorks[2], Joseph A. Finlon[1], Edward P. Nowottnick[2,3], Athena A. Floutsi[4], and David P. Donovan[5]

[1]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740, USA
[2]NASA Goddard Space Flight Center, Greenbelt, MD 2077, USA
[3]NASA Headquarters, Washington DC, USA
[4]Leibniz Institute for Tropospheric Research (TROPOS), Leipzig, Germany
[5]Royal Netherlands Meteorological Institute (KNMI), de Bilt, the Netherlands

Correspondence to: Natalia Roldán-Henao (nroldanh@umd.edu)

Abstract. The Goddard Lidar Observation and Validation Experiment (GLOVE) was a field campaign conducted from 27 January to 28 February 2025, at the NASA Armstrong Flight Research Center, situated at Edwards Air Force Base in California. One main objective of GLOVE was to validate atmospheric data products, including clouds and precipitation, obtained from the Earth Cloud, Aerosol, and Radiation Explorer (EarthCARE) satellite mission. Using the NASA high-altitude ER-2 aircraft, GLOVE deployed the Cloud Physics Lidar (CPL) to capture vertical profiles of aerosols and cloud features along EarthCARE's flight paths. This study validates the EarthCARE baselines BA and BD Atmospheric Lidar (ATLID) Level 2A Target Classification (A-TC) and Cloud Top Height (A-CTH) products against CPL observations from GLOVE. Results show substantial agreement between ATLID and CPL, with a better alignment under homogeneous than heterogeneous conditions, likely driven by ATLID's spatial averaging and multiple scattering effects. In addition, ATLID detects a higher fraction of cloud features than CPL, especially in baseline BA, with a mean probability of detection of 0.85, though at the expense of a higher false-positive rate, misclassifying about 7.6% of clear-sky pixels as cloudy. Baseline BD has a smaller probability of detection (0.74) but reduces the false-positive rate to 4.8%. Overall, cloud top height retrievals show good agreement with CPL. A-TC provides broader sample coverage with relatively good performance, while A-CTH retrieves fewer samples due to a stricter detection threshold, but achieves lower absolute errors and higher correlations. Results for baselines BA and BD show that the majority of samples in both A-TC and A-CTH meet the ATLID cloud top error requirement, with at least 76% of cloud top estimates falling below the 0.3 km error threshold. The findings here highlight the role of spatial averaging, multiple scattering, and cloud heterogeneity in ATLID retrievals, factors that should be considered when using ATLID cloud products.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

1 Introduction

The successful launch of the joint European Space Agency (ESA) and Japan Aerospace Exploration Agency (JAXA) Earth Clouds, Aerosol, and Radiation Explorer (EarthCARE) satellite mission (Wehr et al., 2023) in May 2024 has provided the scientific community with an unprecedented and comprehensive view of atmospheric components. The EarthCARE mission payload includes four instruments: the ATmospheric LIDar (ATLID), the Cloud Profiling Radar (CPR), the Multi-Spectral Imager (MSI), and the Broadband Radiometer (BBR). Together, these instruments deliver global vertical profiles of clouds, aerosols, precipitation, and their associated radiative properties. The resulting measurements have a wide range of applications in climate and weather research (Eisinger et al., 2024; Wehr et al., 2023). For example, they can be used to validate the representation of atmospheric features in models; explore cloud–aerosol interactions and their impact on Earth's energy balance; monitor the vertical distributions of atmospheric liquid water and ice on a global scale, including their transport by clouds and their radiative effects; and analyze the vertical and spatial distribution of the optical properties of aerosols and clouds.

While EarthCARE's payload provides unprecedented measurements of atmospheric features, these new datasets require a comprehensive validation to quantify their uncertainties and biases. To this end, the Goddard Space Flight Center conducted the Lidar Observation and Validation Experiment (GLOVE) field campaign (Yorks et al., 2026) from 27 January to 28 February 2025. This campaign featured the NASA ER-2 platform flying from the Edwards Air Force Base in California. The primary goal of GLOVE was to validate both the EarthCARE and the Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2) missions. The campaign utilized NASA's high-altitude Earth Resources-2 (ER-2) aircraft, which was equipped with four remote sensing instruments, including the Cloud Physics Lidar (CPL) (McGill et al., 2002), the Roscoe Nadir/Zenith Lidar, the Cloud Radar System (CRS) (Li et al., 2004; Walker Mclinden et al., 2021), and the Enhanced MODIS Airborne Simulator (eMAS) (Ellis et al., 2011). Over the course of the campaign, GLOVE completed eight flights and successfully captured six underflight segments of the EarthCARE mission, sampling a range of atmospheric features such as aerosols, cirrus clouds, and stratocumulus clouds.

Among the datasets collected during GLOVE, the CPL observations provide a useful basis for comparison with the EarthCARE's ATLID instrument, although with some limitations. CPL is an elastic backscatter lidar operating at 355, 532, and 1064 nm, with depolarization resolved at 1064 nm. Cloud spatial and optical properties are retrieved using the 1064 and 532 nm channels. Mounted on the ER-2 aircraft (flying approximately at 20 km), CPL provides measurements comparable to satellite observations. These measurements have a higher signal-to-noise ratio (SNR), finer horizontal and vertical resolution, and lower levels of multiple scattering than those from spaceborne instruments, characteristics that make CPL ideal for validating spaceborne lidar instruments (McGill et al., 2007; Yorks et al., 2011b). ATLID, on the other hand, is a high-spectral-resolution lidar (HSRL), linearly polarized with three channels operating at 355 nm (de Villele et al., 2025). These differing

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

wavelengths and measurements principles have only a minor influence on cloud-edge detection, making CPL data ideal to validate the geometrical properties of cloud layers detected by ATLID. Therefore, this investigation focuses on validating the cloud spatial properties (cloud top) of cloud features retrieved by the ATLID instrument.

This paper is organized as follows. Section 2 describes the data and methodology used to validate ATLID cloud spatial properties, including a description of the similarities and differences between CPL and ATLID, an overview of their respective detection algorithms, and a description of the analyzed cases and coincident measurements. Section 3 presents a case study of both homogeneous and heterogeneous clouds, discussing their main sources of difference, as well as the overall mean statistics across the six EarthCARE underflight cloud cases recorded during the GLOVE field campaign. Finally, Section 4 concludes the paper.

2 Data and methodology

2.1 The ATmospheric LIDar (ATLID) Level 2A data products

Operating as a polarized HSRL, ATLID provides direct, simultaneous measurements of backscatter, extinction, and linear depolarization ratio from space. These measurements allow ATLID to perform independent feature masking (Van Zadelhoff et al., 2023) and target classification, including clouds/aerosol discrimination and aerosol layer typing (Irbah et al., 2023). The feature mask and target classification procedures impact the results of multiple Level 2A products, including those generated by the ATLID profile processor (A-PRO; Donovan et al., 2024) and the ATLID layer product processor (A-LAY; Wandinger et al., 2023). In particular, this investigation focuses on two cloud retrieval products. The first is the Target Classification (A-TC; Irbah et al., 2023) product, which classifies detected targets as being e.g. ice clouds, water clouds, and distinct aerosol types, and is used here to retrieve cloud area percentages and cloud top height (CTH). The second is the ATLID Cloud Top Height (A-CTH) product, which provides direct estimates of cloud top features. An overview of the retrieval algorithms for both products is provided below.

During the active EarthCARE mission, several versions of the instruments' products have been released. These versions are called baselines and are named using two capital letters starting with baseline "AA". When minor changes were included, the second letter was changed (i.e., baseline AB, AC, AD). When major updates and software changes occurred, the first letter was changed (i.e., baseline BA, BB, BC). ATLID Level 2A data were publicly released with baseline AE in March 2025. Later in July 2025, baseline BA was released. With baseline BA, the entire dataset was reprocessed, providing one common set of data products available from August 2024 to September 2025. A detailed description of the changes between the different baselines can be found in the "EarthCARE A-PRO Level 2A Processor Disclaimer" for the A-TC product and the "EarthCARE A-LAY Level 2A Processor Disclaimer" for the A-CTH product; both disclaimers are part of the "EarthCARE Product Data

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

Handbook" (European Space Agency, 2026). Since the quality of the products is influenced by the baseline, in this work we use and compare baselines BA and BD.

As mentioned, geophysical properties are retrieved by two processors, A-PRO and A-LAY. The first Level 2A processor, A-PRO, relies on the ATLID Level 1 data (A-NOM), meteorological data (X-MET), the joint standard grid (X-JSG; which is used to guide the binning of the products across EarthCARE's four instruments to facilitate instrument and data synergy), and the outcome of the feature mask product (A-FM; van Zadelhoff et al., 2023) to compute aerosol and cloud optical properties as well as atmospheric feature classification. A-PRO derived geophysical properties distributed across four different output products, namely the A-AER (ATLID-AERosol) product, the A-EBD (Atlid Extinction, Backscatter and Depolarization ratio) product, the A-TC product, and the A-ICE (Ice cloud properties) product. The accurate retrieval of these geophysical quantities requires a high signal-to-noise ratio (SNR), which is usually achieved by averaging multiple along-track profiles. The main problem is that ambiguous or non-physical results can arise when averaging weak and strong signals indiscriminately. To solve this problem, the A-FM product provides A-PRO with a feature detection mask generated using image reconstruction techniques (van Zadelhoff et al., 2023). This feature mask serves as a guide to average the signals and avoid non-physical results in the A-PRO processor.

The A-PRO processing chain begins with the A-AER procedure. This procedure uses direct HSRL retrieval methods to determine extinction and backscatter on large horizontal scales (more than 50 km). To achieve this, the lidar signal is masked and smoothed, guided by the A-FM product and lidar scattering ratios. Strong scattering features, like clouds, are horizontally smoothed less than weaker more homogeneous scattering signals, like those corresponding from aerosols fields. A-PRO then applies a traditional HSRL approach to compute particle extinction, backscatter coefficient, lidar ratio, and depolarization ratio for aerosol and clouds. This procedure ends with a simple multiple scattering correction.

Once the A-AER product is completed, A-PRO retrieves particle extinction and backscatter profiles at a 1 km horizontal scale using the A-EBD procedure. At this finer spatial scale, the molecular SNR is too low to enable the techniques employed by the A-AER procedure. Instead, A-EBD employs a column-by-column forward modelling optimal estimation (OE) approach, using the lidar ratio estimates and aerosol classification generated by A-AER as inputs. Given the lower SNR at these scales, A-EBD operates on a layer-by-layer basis, meaning that the retrieved lidar ratio is assumed to be constant for each aerosol or cloud layer. This procedure accounts for a lidar multiple scattering correction in a manner which is more accurate than the one applied in A-AER.

The extinction, backscatter, and depolarization ratios retrieved in the previous steps are then used to generate the A-TC product. In addition to these lidar-derived quantities, A-TC also requires meteorological data (e.g. temperature and tropopause height) to classify targets into classes such as water or ice clouds. The A-TC procedure follows a decision tree structure in which

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

various conditional thresholds are applied (Irbah et al., 2023). A-TC produces classes at three spatial resolutions, low, medium, and high. However, no significant differences were found between these three resolutions for the cloud features used here. Therefore, results are only shown for the high spatial resolution at 1 km.

Besides A-TC, this investigation also covers the A-CTH product. A-CTH is produced by the A-LAY processor, downstream of A-PRO in the Level 2A processing chain. A-LAY uses the ATLID Mie co-polar signals (A-NOM), the X-MET, the X-JSG, and the Earth's surface elevation from the A-TC product as inputs to retrieve CTH and optical and geometrical properties of aerosol layers. Within this processor, A-CTH is retrieved by applying a Wavelet Covariance Transform (WCT) to the normalized Mie co-polar signal, using a configurable dilation factor that changes depending on the baseline (Wandinger et al., 2023). For baseline BA the dilation factor was set to 6 pixels (~ 600 meters) and for baseline BD to 2 pixels (~ 200 meters). Strong contrasts in the lidar signal, such as at the interface between clouds and clear-sky regions, yield high WCT values. The retrieval algorithm therefore searches for the uppermost WCT maximum between the surface and the tropopause height. If this maximum exceeds a given threshold, it is defined as a cloud top. This methodology is applied at two spatial resolutions (1 X-JSG pixel —about 1 km horizontal resolution— for strong features and 11 X-JSG pixels for weak features) and extends to multiple cloud layers through additional search loops. Quality control flags are additionally provided by comparing the retrieved cloud tops with the A-TC product. In this paper, we used the baselines BA and BD A-CTH products at 1 km resolution (European Space Agency, 2026).

2.2 CPL level 2 data products

Cloud and aerosol properties measured by CPL are provided in the Level 2 Profile and Layer products. The CPL data processing algorithm retrieves physical and optical properties of cloud and aerosol layers using a threshold-based layer detection method based on the 532 nm attenuated scattering ratio (CPL attenuated total backscatter divided by a modeled attenuated molecular backscatter) and the 1064 nm depolarization ratio (Yorks et al., 2021). The threshold profile is computed from range-independent (e.g., dark counts) and range-dependent sources of noise (shot noise from molecular backscatter) (Yorks et al., 2011b), and a layer is detected when the scattering ratio is greater than the threshold profile.

Once atmospheric layers have been detected, the cloud-aerosol discrimination (CAD) algorithm separates them as aerosols, clouds, and undetermined features using a threshold-based technique followed by a multidimensional probability density function (PDF) technique (Yorks et al., 2021). In the threshold-based technique, the algorithm uses the layer-integrated attenuated backscatter (γ') at 532 nm, the layer-integrated 1064 nm volume depolarization ratio (δ'), the layer-integrated 1064 to 532 nm color ratio (χ'), and the midlayer temperature (T_{mid}) to classify the layer by comparing these values with predetermined thresholds. When a layer doesn't meet any criterion in the threshold-based method, then the PDF technique is used. In this technique, the probability of a layer being classified as aerosol or cloud is computed based on a feature type score computed using the following equations.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

$$CAD = \frac{\left(\frac{P^{PDF1}}{100}\right)T^{PDF1} + \left(\frac{P^{PDF2}}{100}\right)T^{PDF2} + \left(\frac{P^{PDF3}}{100}\right)T^{PDF3}}{T^{PDF1} + T^{PDF2} + T^{PDF3}} \times 10 \quad (1)$$

$$P^{PDFi} = \frac{P^{PDFi}c - P^{PDFi}a}{P^{PDFi}c + P^{PDFi}a} \times 100 \quad (2)$$

where i = 1, 2, and 3 correspond to each one of the measured attributes (γ', δ', χ'), $P^{PDF}a$ and $P^{PDF}c$ denote the aerosol and cloud probabilities, respectively, and T^{PDFi} is the total number of data points for the specific attribute i. This feature type score varies from −10 to 10 for each atmospheric layer, and its sign identifies a layer as either cloud (positive) or aerosol (negative). The magnitude of the score represents the confidence in the classification.

2.3 Differences between CPL and ATLID

There are several fundamental differences between ATLID and CPL that impact the retrieved cloud features of the two instruments. The first and most significant one is the different wavelength and measurement principle. As mentioned in the introduction, ATLID is an HSRL operating at 355 nm, while CPL is an elastic backscatter lidar with three wavelengths, which uses the 532 and 1064 nm channels to retrieve cloud properties. Thus, an "apples-to-apples" comparison of the attenuated backscatter data is not possible as the CPL assumes a molecular backscatter signal using external meteorological data to separate molecular and particulate signals, while ATLID is capable of directly measuring them. This difference, however, has a small impact on the cloud-edge detection, due to the strong return signal of cloud layers compared to clear sky regions, making cloud boundaries readily distinguishable for both lidar instruments. For example, Silber et al. (2018) compared the liquid cloud base height (LCBH) retrieved by an HSRL and a Micropulse Lidar (MPL) and found a difference of approximately 20 meters.

In addition to the different measuring principles, ATLID and CPL also differ in several hardware specifications. For example, the ATLID transmitter emits high-energy pulses of 38 mJ with a low repetition rate of 51 Hz, which corresponds to about 140 m spatial resolution. Typically, two pulses are averaged on board giving the L1 attenuated backscatter signals an effective resolution of 280 m. Further, L2 products are reported at the X_JSG resolution (about 1 km) (Illingworth et al., 2015). These specifications contrast with CPL, which emits low-energy pulses of 50 μJ at 1064 nm with a high repetition rate of 5000 Hz. CPL does not record every profile but integrates them to 10 Hz before saving (500 shots accumulated), yielding a raw spatial resolution of approximately 20 m given the ER-2 speed of 200 m s^{-1}. For the final product, CPL data are reported at 1 Hz (approximately 200 m spatial resolution) by averaging 10 profiles at each time increment (McGill et al., 2003). Table 1

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

summarizes the main hardware specifications of ATLID and CPL. ATLID, onboard EarthCARE's platform, has an orbital height of 393 km, much higher than the typical ER-2 altitude of 20 km, resulting in a greater distance from atmospheric features. These differences lead to notable variations in the SNR, with CPL exhibiting higher SNR, and different footprint sizes for ATLID (≤ 30 m) and CPL (1 m). The small footprint of CPL also reduces multiple scattering effects compared to ATLID. As a result, CPL is well-suited for the validation of ATLID cloud geometrical properties. In addition to differences resulting from hardware specifications, ATLID and CPL differences could also steam from the different detection thresholds and algorithms discussed in the previous section.

Table 1 Fundamental differences between CPL and ATLID instruments

Parameter	**CPL**	**ATLID**
Laser Repetition Rate	5.0 kHz	51 Hz
Laser Pulse Energy	25 μJ (532 nm) and 50 μJ (1064 nm)	38 mJ
Receiver FOV	100 μrad	64 μrad
Receiver Footprint Diameter	1 meter at 10 km	≤ 30 m
Detection method	photon counting	analog
Vertical Resolution Sampling Interval	30 m	103 m (up to 20 km) and 500 meters from 20-40 km
Along track sampling interval	200 m (500 shots accumulated at a raw resolution of 10 Hz and then average to 1 second)	285 m (2 profiles accumulated onboard)
Depolarization channel	1064 nm	355 nm

2.3 Coincident measurements

The GLOVE field campaign conducted eight flights, including six EarthCARE underflight segments (Yorks et al., 2026). These flights were performed over both marine and continental regions and captured a range of cloud, aerosol, and precipitation features. Figure 1 shows the flight tracks and the EarthCARE location for each date and Table 2 summarizes the latitude, longitude, overpass time, and the ground track offset at the exact coincidence time of the collocated measurements. Among these segments, a total of nine cloud cases were selected to validate ATLID's cloud geometrical properties. The selection was based on two criteria: (1) the cloud features were observed by both instruments, and (2) the features were recorded by CPL within a 20-minute window centered on the overpass time.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

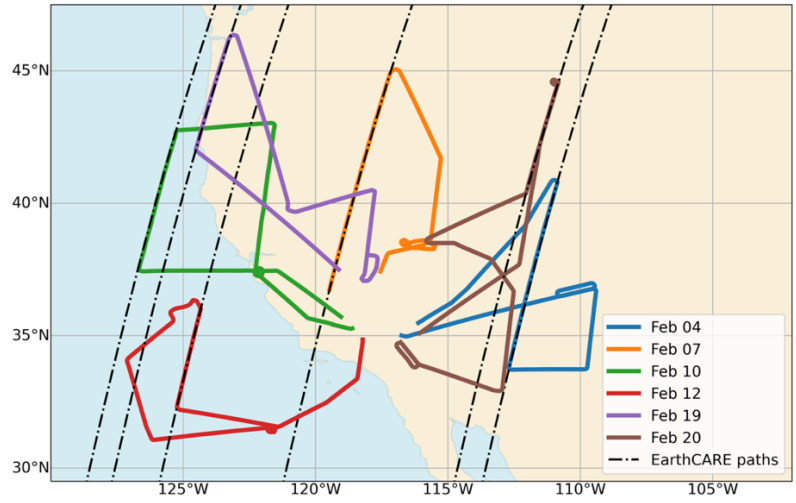


Figure 1: Field track for the six EarthCARE underflight cases during the GLOVE field campaign.

Table 2 Collocation Details of CPL and ATLID Measurements. Offset represents the minimum distance between the satellite and the aircraft around the overpass time.

Flight date (2025)	Latitude range	Longitude range	Overpass time (UTC)	Offset (m)	Description cloud cases
4 Feb	33.7 to 37.8 N	112.7 to 111.7 W	21:47:32	2463	Cirrus cloud observed over a continental region.
7 Feb	41.0 to 45.1 N	118.4 to 116.9 W	22:15:38	1230	Persistent ice cloud observed below 7 km altitude.
10 Feb	37.9 to 42.3 N	126.6 to 125.4 W	22:46:08	2425	Cirrus clouds possibly embedded within an elevate d aerosol layer.
12 Feb	32.1 to 36.3 N	125.2 to 124.3 W	22:36:36	398	Marine stratiform water cloud observed, with an ice cloud layer above.
19 Feb	42.4 to 46.4 N	124.4 to 122.9 W	22:40:57	1707	Marine stratiform water cloud observed, with two cirrus cloud layers above.
20 Feb	39.9 to 43.8 N	113.1 to 111.1 W	21:49:56	561	Persistent ice cloud observed above 7.5 km altitude .

As mentioned, spaceborne lidars such as ATLID, have a lower SNR than aircraft-based observations due to the greater distance from atmospheric features. A common practice to enhance SNR in lidar systems is to spatially average multiple profiles. In the case of ATLID, this averaging is guided by the A-FM product, which provides a feature detection mask based on several image reconstruction techniques, including a hybrid median filter for strong features (Russ, 2007) and a data-smoothing strategy based on a simplified maximum entropy method (Smith and Grandy, 1985) for weaker features. This methodology improves upon previous methodologies by better preserving the cloud–aerosol structure. However, it may still lead to mixing between cloud and clear-sky features, primarily in the presence of weak or broken clouds. Therefore, in this investigation, we present one example of a homogeneous and a heterogeneous cloud feature. In total, three cloud cases were classified as

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

heterogeneous and six cases as homogeneous. Table 2 provides a description of the six underflight segments as well as their observed cloud cases.

Two primary caveats arise when comparing ATLID and CPL. First, imperfect collocation between the aircraft and the satellite means that the instruments do not sample the exact same cloud scene. As shown in Table 2, the ground track offset ranges from 398 m to 2463 m, with an average of 1464 m, making the assumption of horizontal homogeneity necessary. Second, the large difference in platform speeds introduces an additional constraint: EarthCARE covers 4° in approximately 62 seconds, whereas the ER-2 requires about 40 minutes to cross the same distance. This temporal mismatch increases the uncertainties in the assumption of horizontal homogeneity in the coincidence segment, especially at distances further away from the overpass time. Therefore, this study limits the analyses to ±10 min CPL time of the exact coincidence point, which results in a segment of approximately 2° in latitude. This timeframe provides a sufficient sample for validation across the six underflight segments while reducing uncertainties associated with the horizontal homogeneity assumption.

3 Results

3.1 Case study homogeneous cloud: 4 February 2025

Figure 2 presents the total attenuated backscatter (ATB) measured by the level 2 profile CPL (532 nm) product and the 355nm ATLID total ATB (formed by summing the component molecular, particulate and cross-polarized L1 ATBs and gridding to the X-JSG) baseline BA ATLID_EBD_2A product on 4 February 2025. On this date, both instruments detected a geometrically thin, homogeneous ice cloud over Northern Arizona. The red horizontal line indicates the coincident time at 21:47 UTC (Table 2), and the figure encompasses a ±10-minute interval around this point. Qualitative comparison reveals strong agreement between the ATB measurements from ATLID and CPL, with both instruments identifying a similar cloud feature at an approximate mid-layer altitude of 7.5 km. The primary difference is observed in the northern extent of the cloud, where the CPL measurement extends to latitude 36.5°N, while ATLID extends to approximately 36.3°N. This discrepancy may be attributed to the instruments observing different scenes in this region, given the rapid variability of cloud features and horizontal offset of 2.4 km.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

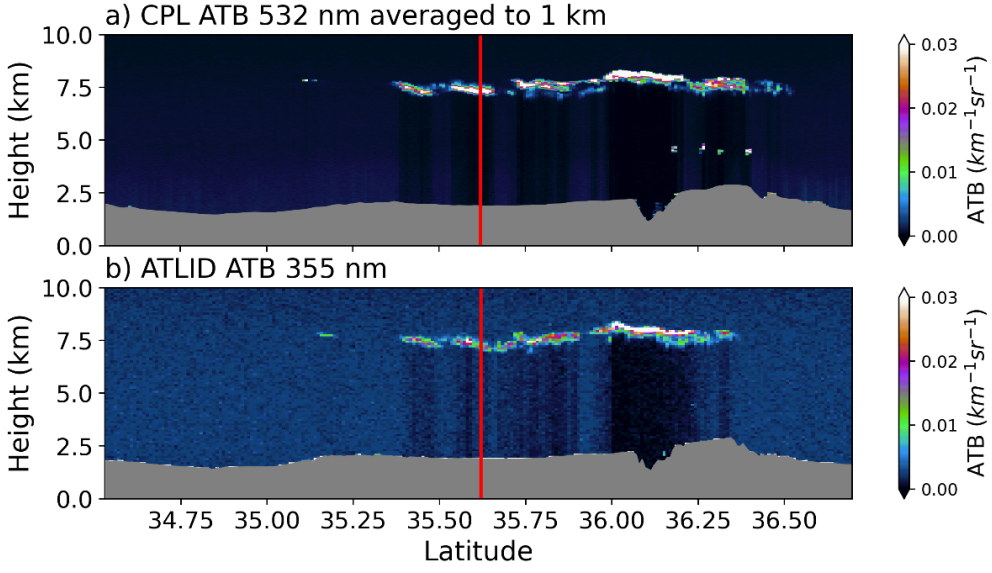


Figure 2: (a) CPL total attenuated backscatter (ATB, km^{-1} sr^{-1}) averaged to 1 km resolution for the 4 February 2025 case. The red vertical line indicates the coincident time. (b) Same as panel (a), but for the ATLID baseline BA A-EBD product. Note that CPL ATB is shown at 532 nm, while ATLID ATB is at 355 nm.

Figure 3 presents a comparison of feature type classifications retrieved by the level 2 profile CPL product and the ATLID A-TC product for baselines BA (panel b) and BD (panel c) on the same date as Fig. 2. The homogeneous ice cloud is depicted in purple. Since this investigation focuses on cloud features, all aerosol types are consolidated into a single class labeled "aerosol." Stratospheric feature classes included in the ATLID classification product are excluded, as they are not pertinent to this analysis. For CPL, the feature type and cloud phase variables are combined to generate the classification shown. Figure 3 demonstrates notable agreement between the cloud edges retrieved by both instruments, although minor differences are evident. For instance, as indicated in Fig. 2, the CPL cloud feature extends to higher northern latitudes due to differences in the observed scenes. For baseline BA (Fig. 3b), ATLID retrieves a slightly thicker cloud between 36°N and 36.25°N than CPL (Fig. 3a), an interval characterized by strong attenuation of the lidar signal below the cloud base. This thicker cloud is likely attributable to multiple scattering effects, given the differences in receiver footprint diameters (Table 1). This thicker cloud could also result from the statistical image reconstruction used in the A-FM product that is used as an input in A-TC. Interesting, Fig. 3c shows that baseline BD has a geometrically thinner ice feature than baseline BA, which better aligns with CPL cloud thickness.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

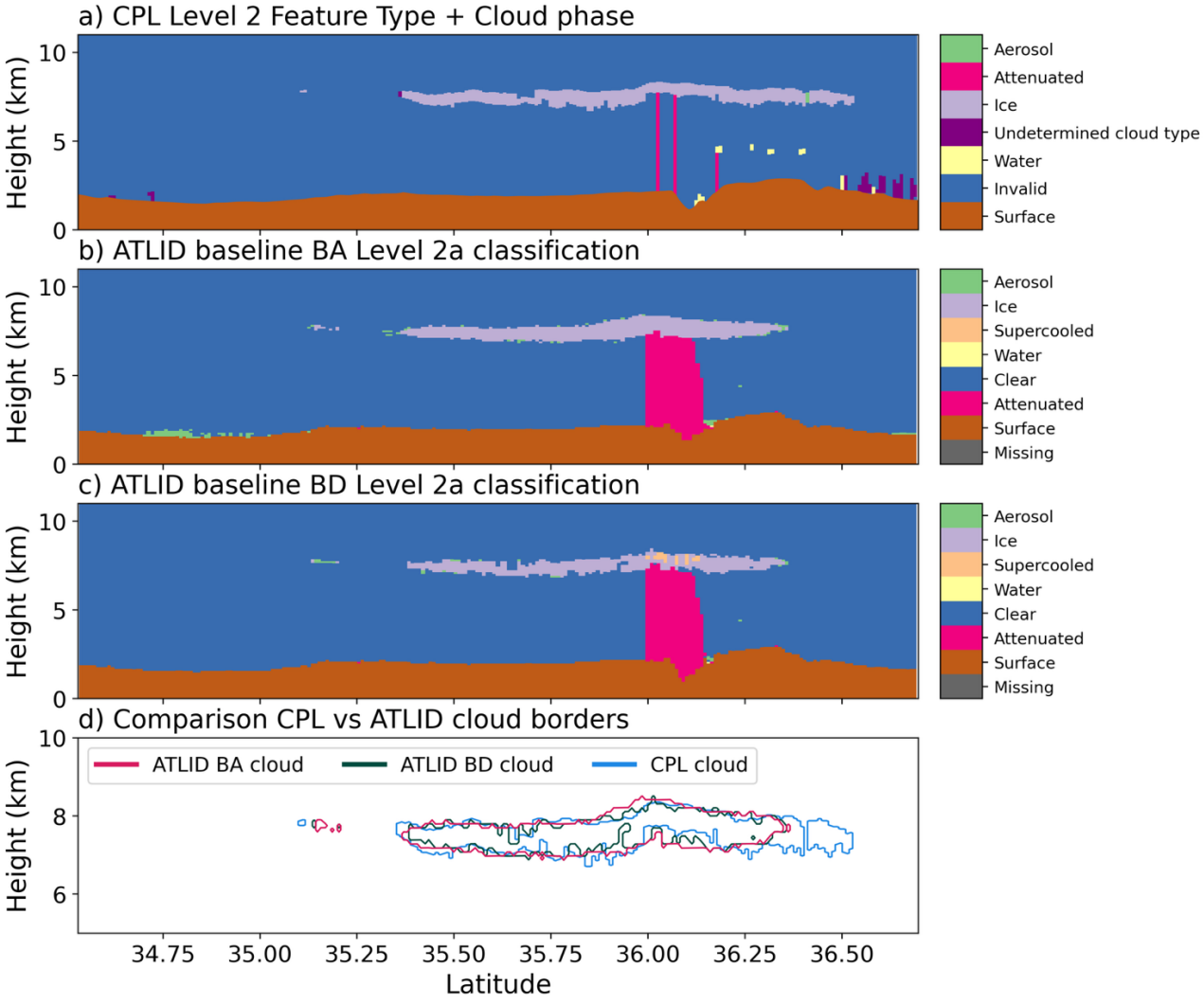


Figure 3: Feature type classification for (a) CPL, (b) ATLID baseline BA, and (c) ATLID baseline BD A-TC product on 4 February 2025. CPL classification is obtained by merging the feature type and cloud phase variables from the CPL Level 2 Profile product. Panel (d) overlays the cloud borders from CPL and the two ATLID baselines to facilitate direct comparison.

Multiple scattering depends on particle properties (number density, size distribution, and shape) and lidar sensing geometry (FOV and distance to the target) (Eloranta, 1998). ATLID and CPL differ not only in their FOV (Table 1) but also on their distance to the target. The greater range of ATLID results in a larger physical footprint at cloud level, capturing multiple-scattered photons in the backscatter direction that would fall outside the footprint of a closer instrument such as CPL. This greater multiple scattering contribution can be significant both within the cloud and in the molecular atmosphere below it, affecting the retrieved extinction and optical depth as well as the spatial extent of the feature. In the case of the spatial extent, multiple scattering effects are more pronounced in cases where the atmospheric layer is highly opaque. Because ATLID

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.



experiences a more significant multiple scattering contribution than CPL, it is able to penetrate deeper into an opaque cloud layer, resulting in a thicker retrieved cloud layer.

Figure 4 provides a more quantitative evaluation of the agreement between the cloud masks of ATLID baselines BA and BD and CPL. Figure 4a compares the vertical profile of the cloud area percentage, computed as the ratio of the number of cloud pixels to the total number of pixels at a given height. As shown, there is a nearly identical vertical profile of the cloud area percentage between ATLID baseline BA and CPL, suggesting this product has an excellent agreement. However, this result is counterintuitive as the cloud feature detected by CPL extends farther north than that of ATLID. To shed some light into this problem, Fig. 4b and 4c show a normalized confusion matrix in which each pixel is classified into one of four categories: true positives (TP, cloud detected by CPL and ATLID), false negatives (FN, cloud detected by CPL but missed by ATLID), false positives (FP, cloud detected by ATLID but not by CPL), and true negatives (TN, clear sky identified by both ATLID and CPL). The matrix is normalized by row to express detection rates rather than raw counts, with raw counts also provided for reference. For baseline BA, the confusion matrix reveals that despite the similar cloud area profiles, 27.2% of CPL cloud pixels are missed by ATLID (FN = 238). This number includes the pixels missed over the northern flank of the ice feature and the low-level water and undetermined clouds detected only by CPL. On the other hand, ATLID baseline BA also detects cloud in regions where CPL observes none (FP = 220), especially over the region between 35.75°N and 36.25°N where ATLID cloud tends to be slightly thicker. For the ice cloud located around 7.5 km, the spatial underestimation in the northern extent of the cloud feature and the spatial overestimation elsewhere approximately cancel out in the vertical profile, producing nearly identical profiles.

For baseline BD, the overestimation of cloudy pixels around the region between 35.75°N and 36.25°N is significantly reduced, leading to a slightly thinner cloud with fewer clear sky pixels incorrectly classified as cloudy (FP = 139). This reduction in the number of false positives makes the underestimation of the cloud feature at the northern extent more apparent. Therefore, the vertical profile of cloud area percentage is smaller than that of CPL and ATLID baseline BA. This smaller vertical profile of cloud area percentage better reflects the differences in the observed scenes between ATLID and CPL.

From the confusion matrix, we computed three main statistics: the Probability of Detection (POD = TP / (TP + FN)), also known as the hit rate, the Critical Success Index (CSI = TP / (TP + FP + FN)), and the Heidke Skill Score (HSS). POD quantifies the fraction of clouds detected by CPL that are also identified by ATLID and ranges from 0 to 1, where 0 indicates that none of the CPL cloud pixels are detected by ATLID and 1 indicates perfect detection of all CPL cloud pixels. CSI measures overall detection agreement while excluding true negatives, thereby avoiding inflation of the metric due to the large number of clear-sky pixels. CSI also ranges from 0 to 1, where 0 indicates no agreement and 1 represents perfect agreement. Finally, HSS measures classification skill relative to random chance. It ranges from $-\infty$ to 1, where negative values indicate performance worse than random chance, 0 indicates no skill, and a value of 1 corresponds to a perfect classification. For

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

baseline BA, POD is 0.72, CSI is 0.58, and HSS is 0.72 when considering a ±10 min time window around the overpass. For baseline BD, POD is 0.65, CSI is 0.56, and HSS is 0.71. For reference, Van Zadelhoff et al. (2023) reported a hit rate of 0.68 and an HSS of 0.74 for the A-FM product validated against a simulated truth, noting that the A-FM includes any type of atmospheric feature, not exclusively clouds. When the window is reduced to ±5 min, the POD of baseline BA (BD) increases to 0.85 (0.75), the CSI to 0.64 (0.63), and the HSS to 0.77 (0.76), indicating that part of the disagreement is attributable to the two instruments sampling different cloud scenes as the temporal separation between observations increases. These numbers reveal two things: the ATLID A-TC product performs well overall for homogeneous clouds compared with real reference data, and reducing false positives in BD comes at the expense of a lower POD. This lower POD, however, isn't necessarily negative, as the high POD in baseline BA was inflated by overestimating cloudy pixels.

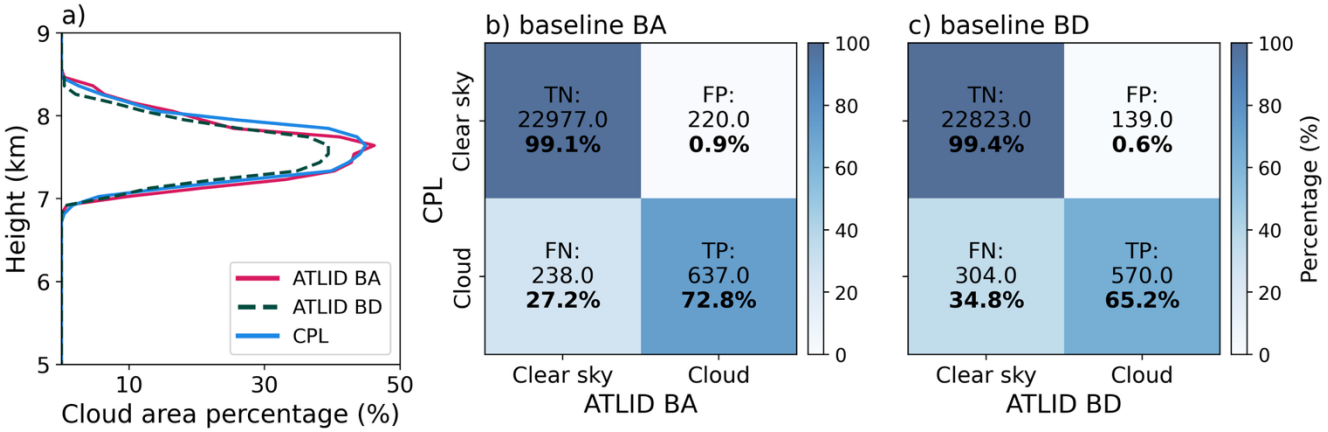


Figure 4: (a) Percentage of cloudy area for ATLID baseline BA (magenta solid line) and baseline BD (green dashed line) and for CPL (blue solid line) for the 4 February 2025 case. (b) Normalized confusion matrix showing true positives (TP; cloud detected by both CPL and ATLID), false negatives (FN; cloud detected by CPL but missed by ATLID), false positives (FP; cloud detected by ATLID but not by CPL), and true negatives (TN; clear sky identified by both instruments) using ATLID baseline BA. (c) same as panel (b) but for ATLID baseline BD. Raw counts and row-normalized percentages are shown.

Comparisons of the estimated CTH from the target classification product (A-TC) and from the A-CTH product are shown in Fig. 5. CTH estimates from A-TC exhibit a strong positive correlation with CPL retrievals (Fig. 5a), with a Pearson correlation of 0.79 (baseline BA) and 0.7 (baseline BD) and a Root Mean Square Error (RMSE) of 0.17 km (baselines BA and BD). Similarly, A-CTH has a correlation of 0.83 (baseline BA) and 0.8 (baseline BD) and a RMSE of 0.2 km (baseline BA) and 0.19 km (baseline BD). One important thing to note is that A-CTH has a reduced number of CTH estimates than the A-TC retrievals, especially in the lower troposphere. Such a reduced sample is likely the result of the strict thresholds used on the SNR and the WCT within this atmospheric region, which reduces the number of retrieved CTHs. As noted, baseline BD shows a smaller correlation than BA. However, this weaker correlation is compensated by a reduction in the absolute error. Figure 5b-d presents the Probability Density Function (PDF) of the absolute CTH errors. For the A-TC product, 88% of errors in baseline BA are below the EarthCARE mission error requirement of 0.3 km (T. Ed. Wehr, 2006), increasing to 92% for BD. Similarly, for A-CTH, 84% of errors in baseline BA fall below the same 0.3 km threshold, increasing to 86% in BD.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

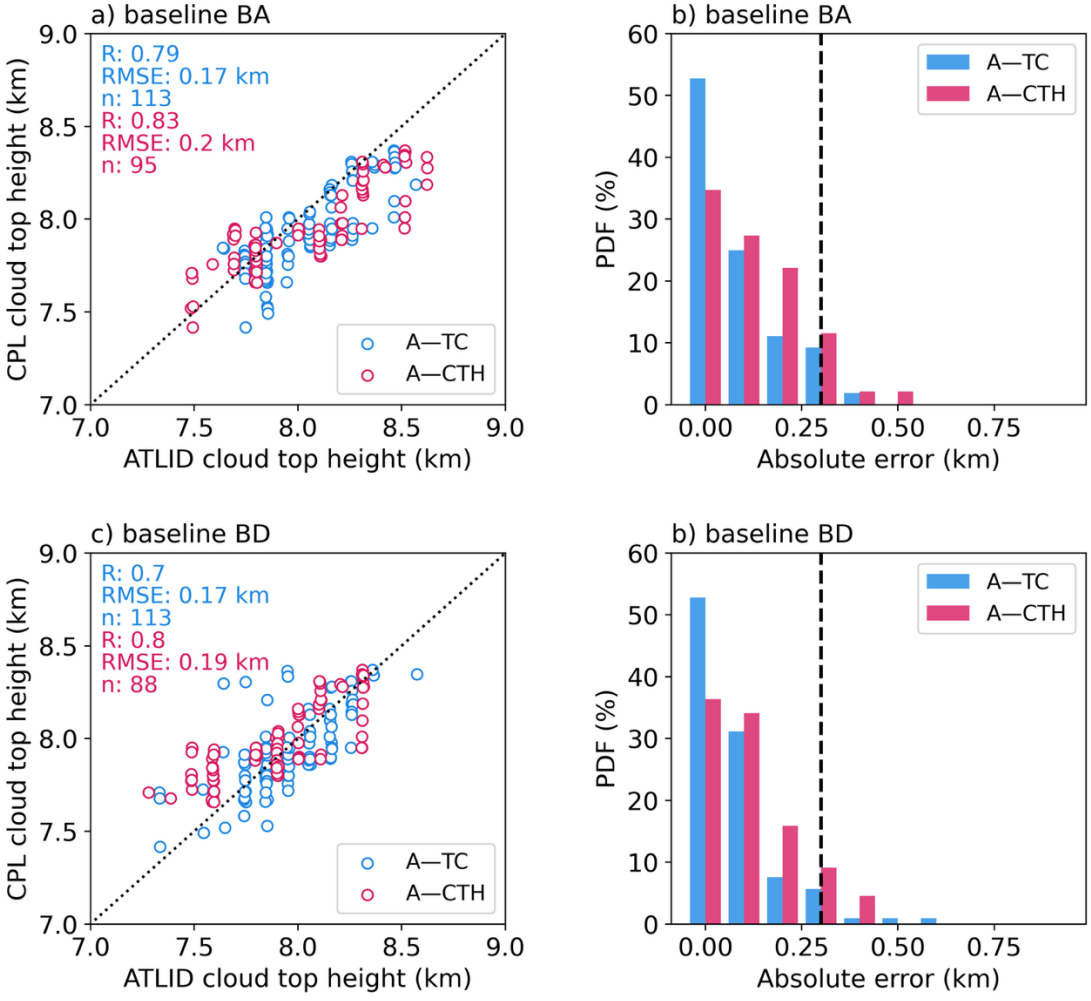


Figure 5: (a) Scatter plot of cloud top height (CTH) from CPL versus the ATLID baseline BA A-TC and A-CTH products during 4 February 2025. R denotes the correlation coefficient, and RMSE denotes the root mean square error. (b) Distribution of absolute CTH error, computed as $|CTH_{ATLID} - CTH_{CPL}|$. (c - d) Same as (a - b) but for baseline BD.

3.2 Case study heterogeneous cloud: 7 February 2025

Figure 6 shows the 355 nm ATLID- and 532 nm CPL-measured ATB for 7 February 2025. On this date, a fragmented ice cloud was observed extending over the state of Oregon. Similar to Fig. 2, the red horizontal line indicates the coincident time at 22:15 UTC (Table 2), and the figure spans ±10 min around this point (2°). From a qualitative perspective, the ATB measurements from ATLID and CPL show a relatively good agreement, with both instruments capturing a similar cloud feature at an approximate mid-layer altitude of 5.2 km. In contrast to the previous homogeneous case, this cloud exhibits a more complex structure, with multiple clear-sky gaps interrupting the main cloud layer and greater variability in cloud-top and cloud-base heights. This complex structure directly influences the feature mask classification and complicates the separation of cloudy regions from clear sky. Figure 7 presents the corresponding feature type classification for the baselines BA and BD

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

A-TC product and for the level 2 CPL feature type product. Unlike the homogeneous case, this case shows a weaker agreement between ATLID and CPL. In the ATLID baseline BA retrieval (Fig. 7a), clear-sky regions between cloudy patches completely disappear, resulting in a cloud field that appears artificially homogeneous. This behaviour observed in the baseline BA A-TC product is likely due to the spatial smoothing and averaging applied in the ATLID Level 2A A-PRO algorithm to enhance the SNR, which leads to smearing of cloud edges and mixing of cloudy and clear-sky signals. This issue is a current point of attention and has been targeted for improvements in the A-FM and A-PRO processor updates. Some of these improvements are already evident in baseline BD (Fig. 7c) which shows a significant reduction in the mixing of cloudy and clear-sky signal.

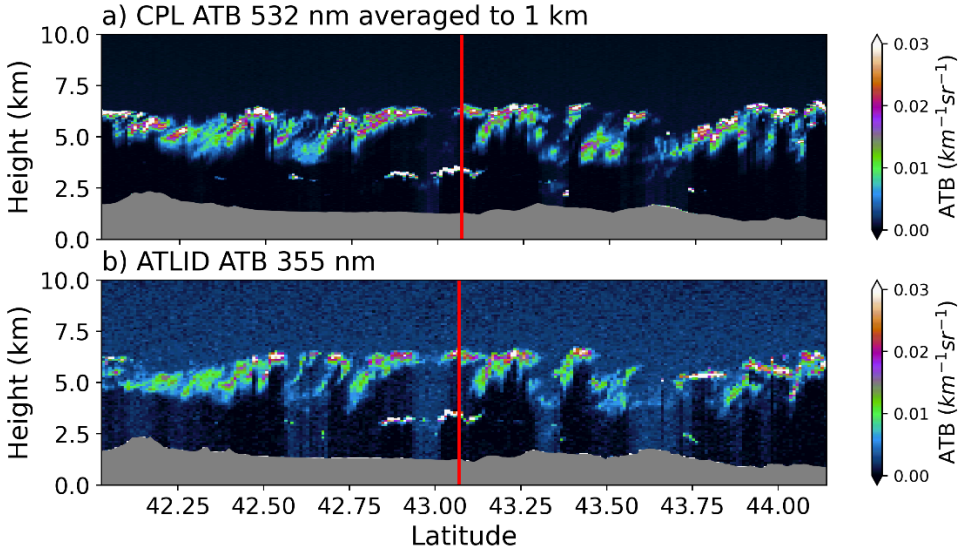


Figure 6: Same as in Figure 2 but for the 7 February 2025 case.

The differences in the cloud mask feature between CPL and ATLID baseline BA and BD are evident in Fig. 7d, which superimposes the cloud edges of the ice features from ATLID products and CPL. In the case of baseline BA, important differences are observed at both the cloud top and base. For instance, the ATLID baseline BA cloud base tends to be lower and more homogeneous than that reported by CPL and ATLID baseline BD, likely due to an enhanced cloud signal resulting from ATLID's averaging process, in addition to the high influence of multiple scattering in the satellite measurements. For the ice cloud tops, regions are observed where the cloud top reported by CPL is higher than that from the two ATLID baselines. This mismatch can be attributed to the differences in the cloud scenes observed by each instrument, given the more dynamic and rapidly evolving nature of this heterogeneous cloud structure and a 1.2 km horizontal offset.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

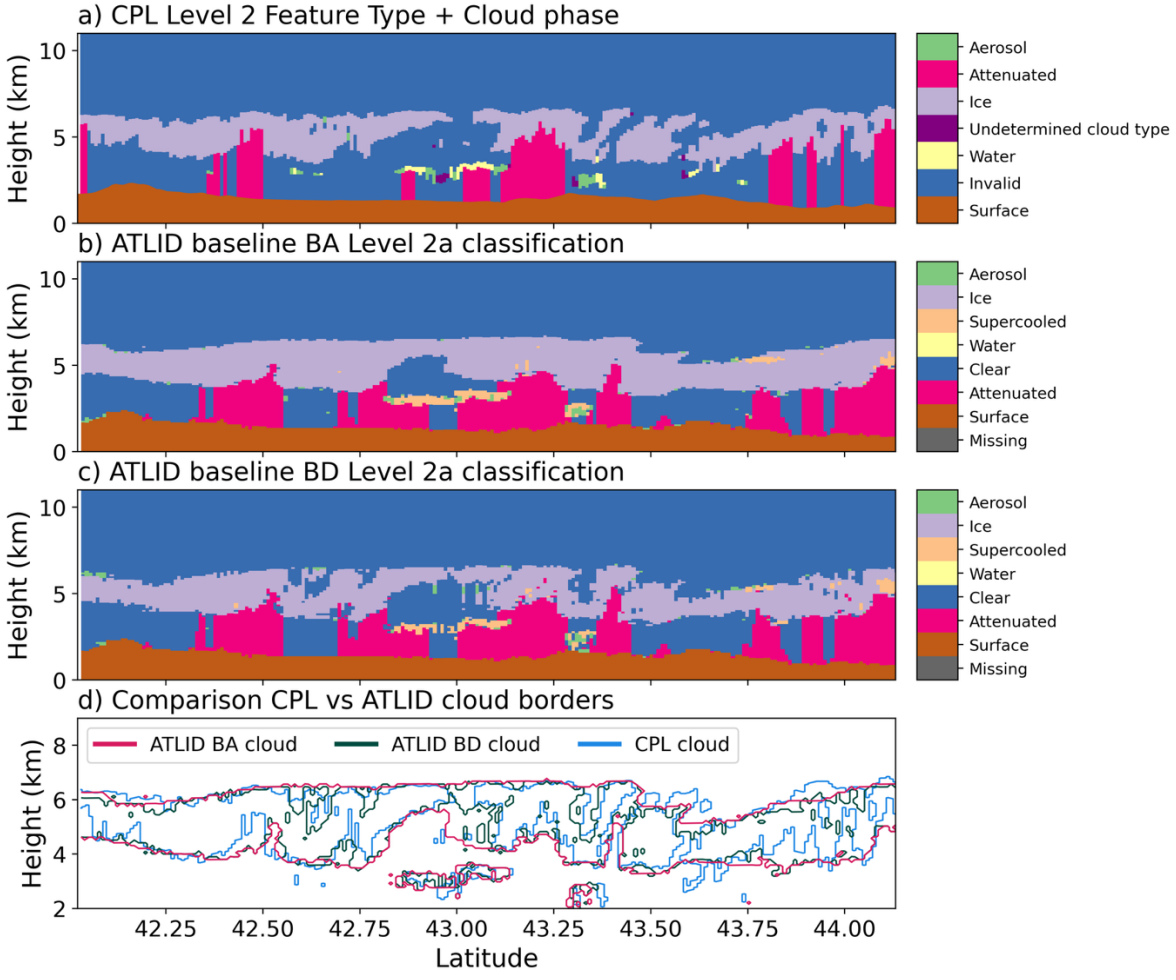


Figure 7: Same as in Figure 3 but for the 7 February 2025 case

Figure 8a presents the quantitative comparison of the cloud area percentage derived from the ATLID A-TC product and the CPL feature mask product on 7 February 2025. For this date, both ATLID baselines report a larger cloud area percentage than CPL, particularly below the CTH. The enhancement of the cloud area is particularly significant for baseline BA, where the cloud area percentage reaches 90% for the ice layer, while CPL reaches a maximum of only about 70%. Contrary to baseline BA, the cloud area percentage of baseline BD peaks at 70%, but this maximum occurs at lower altitudes than CPL. This difference near the cloud top region between ATLID baseline BD and CPL likely reflects the distinct cloud scenes each instrument observes. As shown in Figure 7c, ATLID and CPL show important differences in the observed cloud top between 43.5°N and 44.0°N. These differences at the cloud top are offset in baseline BA by an overfilling of the cloud signal, especially around latitudes 42.5°N and 42.75°N. The substantial reduction in the overfilling in baseline BD makes the differences in the cloud top more apparent, which explains the differences in the vertical profile of cloud area percentage between BA and BD.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

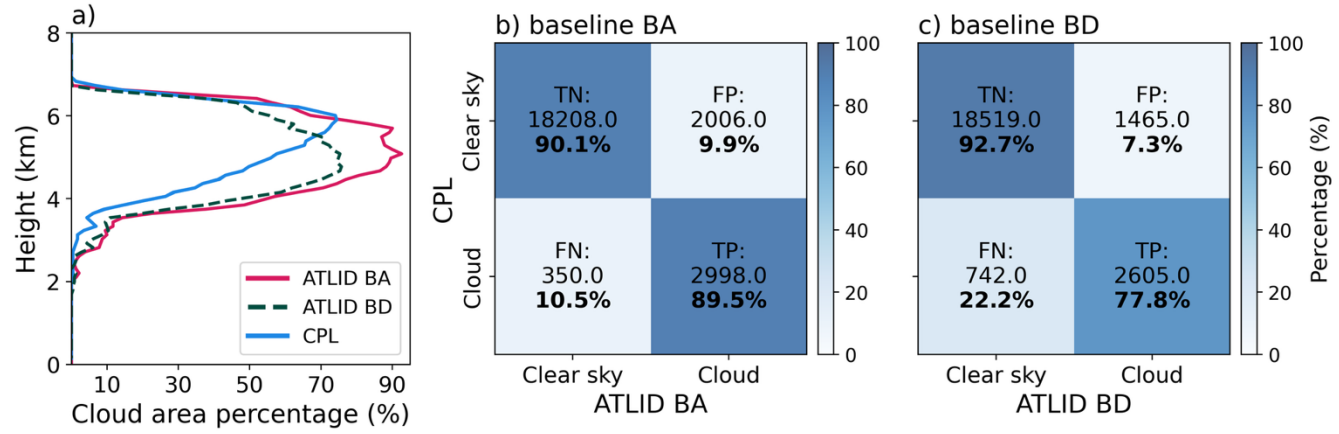


Figure 8: Same as in Figure 4 but for the 7 February 2025 case.

Regardless of the baseline, ATLID seems to overestimate the cloud signal for heterogeneous cloud features. This enhancement of the cloudy signal by ATLID is partly explained by the strong influence of spatial averaging, which mixes cloudy and clear-sky signals as well as a stronger contribution of multiple scattering than the 4 February case. The stronger multiple scattering contribution is consistent with the higher below-cloud attenuation observed in Fig. 6b-c, which indicates a more optically thick layer, conditions under which multiple scattering effects become significant. This is further supported by the confusion matrix in Fig. 7b.c, which shows that 9.9% of the clear-sky pixels are incorrectly classified as cloud by ATLID baseline BA and 7.3% for baseline BD.

The statistical metrics for this date suggest that there might be differences in the performance of ATLID when classifying homogeneous clouds versus heterogeneous ones. POD reaches 0.89 (baseline BA) and 0.78 (baseline BD) for a ±10 min window and 0.94 (baseline BA) and 0.81 (baseline BD) for ±5 min, numbers noticeably higher than the ones reported for homogeneous clouds. This higher POD is expected, since the spatial averaging applied to enhance the SNR inflates the cloud area percentage, making it more likely that a cloudy pixel in CPL is also classified as cloudy in ATLID. This is the same reason why baseline BD has consistently lower POD values than BA. However, this increase in POD comes at the expense of a high number of false positives and classification accuracy, as shown by the CSI and HSS values. For a ±10 min window, CSI drops to 0.56 (baseline BA) and 0.54 (baseline BD), which is 0.02 points lower than in the homogeneous case. HSS decreases to 0.66 (baseline BA) and 0.65 (baseline BD), representing a reduction of 0.06 points from the homogeneous case. The same spatial averaging responsible for the POD increase also raises the number of clear-sky pixels misclassified as cloudy (i.e., the false positives), which explains the corresponding decline in CSI and HSS. In other words, although ATLID detects more of the CPL cloud feature, a larger portion of the scene is misclassified, increasing the overall classification error.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

Figure 9 compares the cloud boundaries retrieved from CPL and ATLID for this heterogeneous case. Regardless of the baseline, both the A-TC and A-CTH products show weaker agreement with CPL CTH than in the homogeneous case, with lower correlations and higher RMSEs. For example, for baseline BA the correlation with CPL CTH is 0.61 for A-TC and 0.4 for A-CTH, compared to 0.79 and 0.83, respectively, for the homogeneous case. Similarly, the RMSE is 0.38 km for A-TC and 0.52 km for A-CTH, compared to 0.17 km and 0.2 km, respectively, for the homogeneous case. Two main conclusions emerge from this analysis: cloud boundary retrievals appear to be less accurate for heterogeneous scenes than for homogeneous ones, and A-TC outperforms A-CTH in this case.

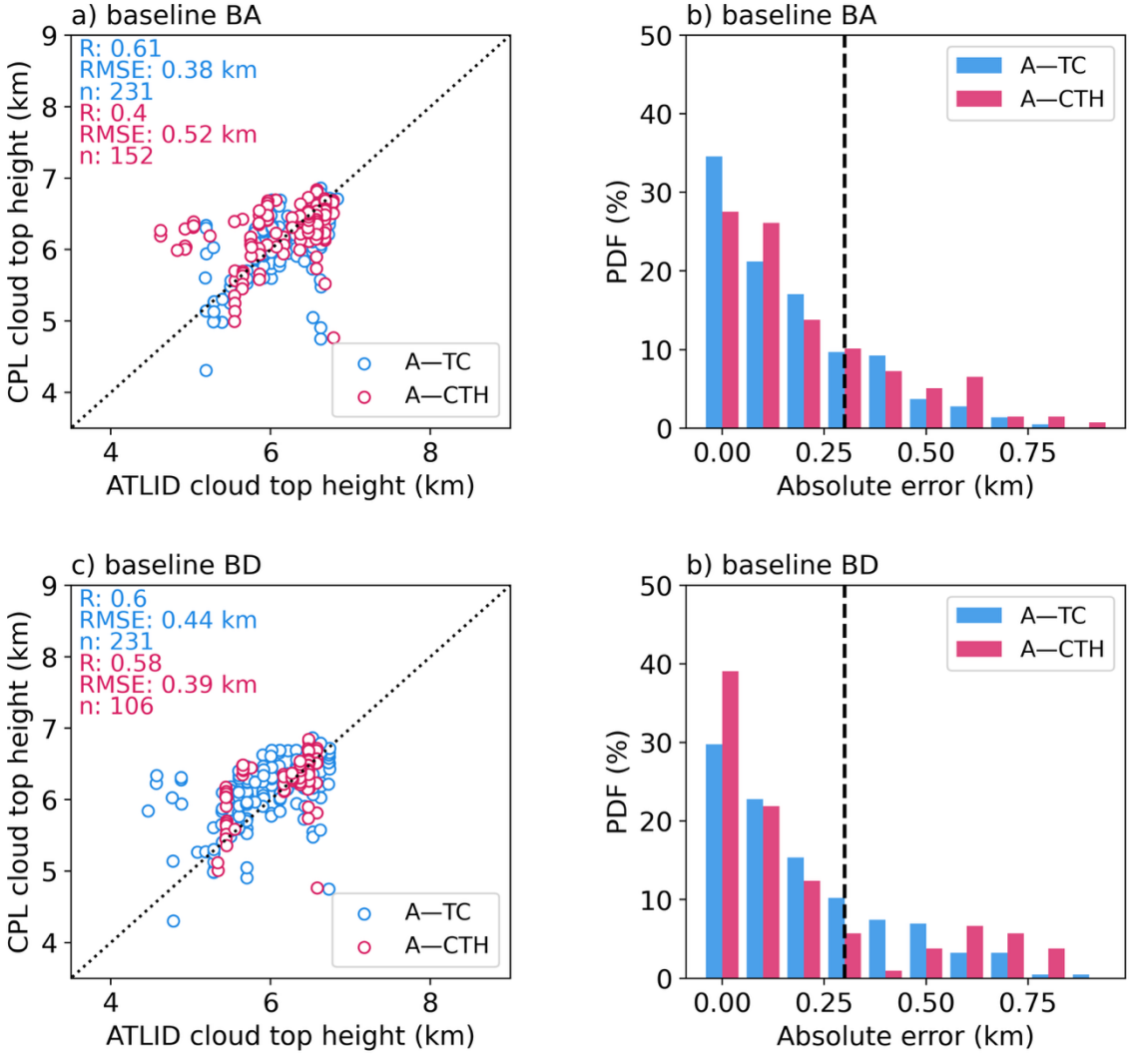


Figure 9 Same as in Figure 5 but for the 7 February 2025 case.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

The weaker agreement in this heterogeneous case likely stems from a few things. First, for this particular case there are some small discrepancies in the cloud phase classification between ATLID and CPL, especially around the top of the cloud feature (see Fig. 7). Since only the ice layer was used to compute the CTH for A-TC, this can lead to differences in the estimated CTH as supercooled water above the ice layer was ignored. Second, the spatial averaging applied by ATLID mixes signals from different cloud structures, smoothing out the sharp cloud top variations typical of heterogeneous fields, an effect that is much weaker in the homogeneous case. Another possible reason for the different performance between the CTHs of the homogeneous and heterogeneous cases is the different meteorological conditions on each day. During 7 February 2025 (heterogeneous case), ATLID measured a tropopause height as low as 7.5 km, right at the top of the ice cloud feature, while during 4 February 2025 (homogeneous case) the tropopause height was close to 17.5 km, a difference of 10 km between the two cases. The tropopause altitude influences the retrieved cloud features, and in the case of the A-CTH product, it affects the altitude-dependent thresholds required to retrieve the CTH. Finally, discrepancies likely also stem from the two instruments sampling slightly different cloud scenes (Table 2). In terms of the different performance between A-TC and A-CTH estimates, it is challenging to determine the reason for such a discrepancy, as A-TC and A-CTH are derived from different methodologies.

Figures 9c and 9d examine the distribution of absolute errors in the CTH estimates for A-TC and A-CTH, respectively. For A-TC, 70% (baseline BA) and 65% (baseline BD) of the estimates have an absolute error below 0.3 km, with an overall mean absolute error (MAE) of 0.26 km (baseline BA) and 0.3 km (baseline BD). For A-CTH, 61% (baseline BA) and 73% (baseline BD) of the estimates have an error below 0.3 km, with an overall MAE of 0.35 km (baseline BA) and 0.26 km (baseline BD).

3.3 Average cloud statistics

Having examined a homogeneous and a heterogeneous case in detail, we now present the average results across all six available EarthCARE underflight segments since individual cases carry limited statistical significance on their own and a more robust assessment of ATLID's performance is needed. Figure 10a shows the average vertical profile of cloud area percentage for ATLID baselines BA and BD and for CPL. Overall, the vertical profile shapes are similar, with both instruments detecting cloud layers at the same altitudes; however, ATLID overestimates the cloud area percentage across the entire profile, reaching a maximum overestimation for baseline BA of approximately 15% around 8 km. This overestimation is substantially reduced in baseline BD, remaining below 10% across the entire profile.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

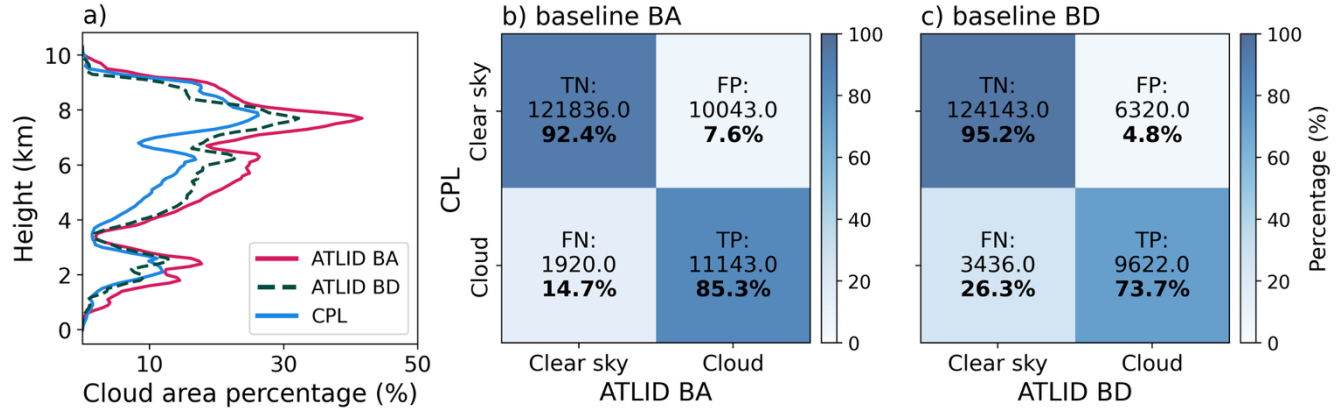


Figure 10 (a) Average percentage of cloudy area for ATLID baseline BA (magenta solid line) and baseline BD (green dashed line) and for CPL (blue solid line) for all six EarthCARE underflights. (b) Normalized confusion matrix showing true positives (TP; cloud detected by both CPL and ATLID), false negatives (FN; cloud detected by CPL but missed by ATLID), false positives (FP; cloud detected by ATLID but not by CPL), and true negatives (TN; clear sky identified by both instruments) using ATLID baseline BA. (c) same as panel (b) but for ATLID baseline BD. Raw counts and row-normalized percentages are shown.

Reducing the overestimation from baseline BA to baseline BD improves the overall skill of the A-TC feature mask. Based on the confusion matrix in Fig. 10b, ATLID baseline BA misclassified 7.6% of clear-sky pixels as cloudy and missed 14.7% of cloudy pixels. For this baseline, the mean POD is 0.85, the CSI is 0.48, and the HSS is 0.61. In contrast, baseline BD incorrectly classifies only 4.8% of clear-sky pixels as cloudy, representing a 2.8% reduction in false positives compared to baseline BA. This reduction corresponds to a 37% decrease in raw counts, with false positives decreasing from 10,043 to 6,320. However, this improvement is accompanied by reduced detection sensitivity, as BD fails to detect 26.3% of cloudy pixels, resulting in a lower POD of 0.74. Despite this, the reduction in false alarms outweighs the decrease in detections, leading to increased overall skill scores, with the CSI rising to 0.50 and the HSS to 0.63.

Regardless of the baseline, the relatively low CSI, compared to the higher POD and HSS, is mainly explained by the number of false positives, which despite the modest false-positive rate, it becomes large in absolute terms because clear-sky pixels vastly outnumber cloudy ones. This reflects an underlying trade-off between cloud detection sensitivity and false-positive rate: ATLID misses few cloudy pixels, but at the expense of some clear-sky pixels misclassified as cloudy. Some of the reasons why ATLID classifies clear-sky pixels as cloudy have been presented throughout the paper, including the smearing of cloud edges and mixing of signals by the spatial averaging, multiple scattering effects, and the strongest attenuation experienced by CPL compared to ATLID. In addition, residual depolarization biases in ATLID, reported during the EarthCARE Science and Validation Workshop 2025 (Kubota et al., 2026), could also contribute since A-TC relies on the depolarization ratio to classify clouds and aerosols. Nevertheless, the results shown here reveal that during daytime conditions EarthCARE detects most cloud features at a fine spatial resolution of 1 km, showing similar locations, shapes and phases as CPL features, while maintaining a low false-positive rate, demonstrating robust discrimination between clear and cloudy conditions. The reduction in false

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

positives from BA to BD indicates that recent baselines are less affected by the mixing of cloud and clear-sky signals, and further improvements are anticipated with future processor releases.

Figure 11 presents the average results for cloud boundaries across all cloud cases. We observed a linear relationship between ATLID products and CPL, with an estimated correlation of around 0.99 and an RMSE of around 0.3 km for the CTH derived from the A-TC and the A-CTH product (with minor differences between baselines), calculated from all cloud cases combined. However, this strong correlation is the result of combining low-, mid-, and high-level clouds that span a wide range of altitudes, which results in a correlation coefficient that is larger than any individual cloud case. Figure 12 shows a box-and-whisker plot summarizing the distribution of individual correlations, RMSEs, and MAEs for each CTH product and baseline. Two main differences stand out. First, the baseline update significantly improved A-CTH correlation, with the median rising from 0.4 in baseline BA to 0.55 in BD. This increase likely reflects better configuration parameters for cloud detection in the lower troposphere, as well as the improved layer-mean depolarization ratio implemented in the newer baseline (European Space Agency, 2026). Second, A-TC has slightly larger errors in the estimated CTH than A-CTH regardless of the baseline. For example, the median RMSE for baseline BA is 0.28 km for A-TC and 0.22 km for A-CTH (see Table 3). However, both products (A-TC and A-CTH) and baselines (BA and BD) have an overall good performance, as it is shown in Table 3, with at least 76% of the samples meeting the EarthCARE CTH accuracy requirement (Wehr, 2006).

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

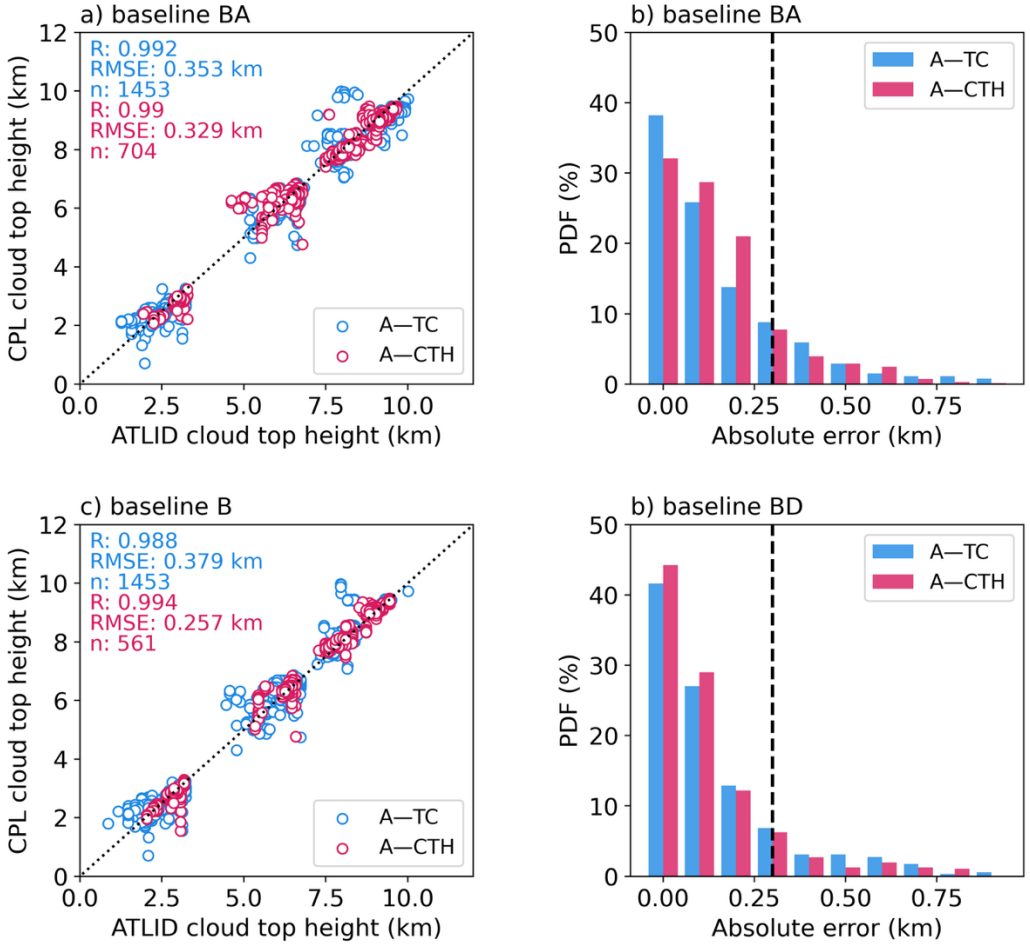


Figure 11 (a) Scatter plot of cloud top height (CTH) from CPL versus the ATLID baseline BA A-TC and A-CTH products across all six EarthCARE underflight cases. R denotes the correlation coefficient, and RMSE denotes the root mean square error. (b) Distribution of absolute CTH error, computed as $|CTH_{ATLID} - CTH_{CPL}|$. (c - d) Same as (a - b) but for baseline BD.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

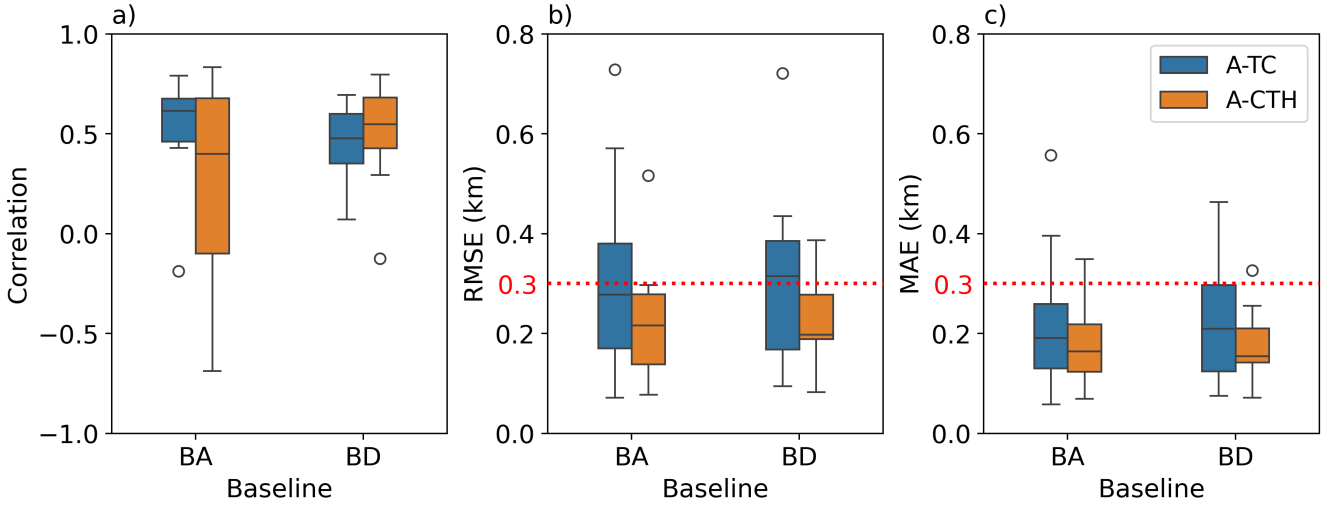


Figure 12 **Box-and-whisker plots summarizing the distribution of (a) correlation coefficients, (b) Root Mean Square Error (RMSE), and (c) Mean Absolute Error (MAE) between the A-TC CTH and A-CTH products against CPL for baselines BA and BD. The central box spans from the 25th to the 75th percentile, with the inner horizontal line marking the 50th percentile (median) The dotted red line highlights the EarthCARE error mission requirement for CTH. Results are for all cloud cases across the six EarthCARE underflight segments.**

The different performance in the correlation and errors between A-TC and A-CTH might be explained by their different retrieval methodologies. A-TC classifies atmospheric features using multiple physical quantities, such as extinction, backscatter, and depolarization ratio, as well as meteorological variables like temperature and tropopause height. This approach, although different, shares similar key principles with the CPL CAD algorithm, which also relies on backscatter, depolarization ratio, and midlayer temperature. These similarities between both methods likely help A-TC and CPL agree more consistently on atmospheric features, preserving the relative variation in CTH across different cloud regimes. Contrary to this, A-CTH identifies CTH solely based on the Mie co-polar signal using the WCT and applying dynamic and configurable detection thresholds that are usually stricter for the lower troposphere. These strict thresholds significantly reduce the number of CTH samples retrieved by A-CTH compared to A-TC (see Fig. 10a), but allows A-CTH to select the clearest, sharpest, most unambiguous cloud-top signals, which can lead to a smaller RMSE. In contrast, A-TC retrieves CTH more frequently, including for weaker and more ambiguous signals that A-CTH's stricter threshold excludes, providing broader and more consistent coverage of the dataset, which is also a desirable feature for cloud boundary retrievals. Table 3 summarizes the main statistics for the CTH retrievals of both baselines (BA and BD) in the A-TC and A-CTH product.

Table 3 Main statistics summarizing the results for different ATLID baselines and products. Median correlation, RMSE, and MAE were computed as the median values of the individual correlations, RMSEs, and MAEs of each cloud case. The percentage of samples with absolute errors lower than 0.3 km is based on the total number of samples.

Product and baseline	Median Correlation	Median RMSE (km)	Median MAE (km)	# samples	% samples with an absolute error lower than 0.3 km

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

A-CTH baseline BA	0.40	0.22	0.16	704	79.69%
A-CTH baseline BD	0.55	0.20	0.15	561	85.03%
A-TC baseline BA	0.62	0.28	0.19	1453	76.15%
A-TC baseline BD	0.48	0.31	0.21	1453	78.20%

4 Conclusions

The launch of the EarthCARE mission in May 2024 has provided the scientific community with a new and unprecedented view of atmospheric components, including cloud and aerosol layers. These new measurements have a wide range of applications in atmospheric science; however, they require a comprehensive validation of their performance. This investigation validates cloud geometrical properties derived from two EarthCARE Level 2A products, namely the ATLID Target Classification product (A-TC) and the ATLID Cloud Top Height product (A-CTH). This evaluation is conducted using six EarthCARE underflight segments taken during the GLOVE field campaign in February 2025. GLOVE used the NASA ER-2 aircraft, based at Edwards Air Force Base in California, to carry four instruments, including CPL, the backscatter lidar used to validate the EarthCARE ATLID instrument.

Results from this investigation reveal a good agreement between the cloud feature mask from ATLID and CPL. Overall, ATLID baseline BA detects 85% of cloudy pixels during daytime conditions at a fine horizontal resolution of 1 km (based on the A-TC product) while this number is 75% for baseline BD. The retrieved clouds display similar shapes, locations, and phases as those from CPL, yielding a similar vertical profile of cloud area percentage between both instruments. This apparent strong cloud detection capability for baseline BA, comes with a modest trade-off, an average of 7.6% of clear-sky pixels misclassified as cloudy, which can be explained by several factors. First, the spatial averaging performed by the A-PRO processor to increase the SNR allows ATLID to detect subtle cloud features that might otherwise be missed, though this process can sometimes mix clear-sky and cloudy signals. This mixing is substantially reduced in baseline BD, with an average of only 4.8% of clear-sky pixels misclassified as cloudy. The reduced percentage of false positives ultimately results in a lower probability of detection for baseline BD, but a more robust agreement between ATLID and CPL feature masks. Results indicate that the mixing of clear-sky and cloudy signal is not uniform across all scenes, being more pronounced for heterogeneous clouds than for homogeneous ones. Analyses show that for the evaluated homogeneous cloud case, the percentage of clear-sky pixels incorrectly classified as cloudy was only 0.9% (baseline BA) and 0.6 (baseline BD), while the same analysis for a heterogeneous cloud case yielded a 9.9% (baseline BA) and 7.1% (baseline BD) overestimation. Another contributing factor to the overestimation of cloudy pixels arises from the stronger multiple scattering experienced by ATLID compared to CPL, which results in thicker clouds, especially when sensing opaque cloud features. Differences also occur because of slight variations in the observed atmospheric scenes, as ATLID on space takes considerably less time than CPL to observe the same

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

region. Finally, CPL can be fully attenuated by overlying features and miss clouds, while ATLID still detects them, a mismatch that reflects a limitation of the reference instrument rather than an error in ATLID's retrieval.

In addition to the feature mask, this investigation also validates the cloud tops retrieved by ATLID. Two estimates were considered: the uppermost pixels of individual cloud features in the A-TC products, and the reported values from A-CTH. Results show a linear relationship for CTH, although correlations vary significantly between individual cloud cases, with a median of 0.62 for baseline BA A-TC CTH and 0.40 for baseline BA A-CTH (0.55 and 0.48 respectively for baseline BD). The lower correlation for A-CTH may arise from several reasons. First, there is an inherent difference in the approach used to retrieve CTH between A-CTH and CPL, while the technique used by A-TC aligns better with CPL. Second, A-CTH has a smaller sample size than A-TC, especially near the Earth's surface. This sample is nearly three times smaller, which could lead to noisier correlations that are more sensitive to outliers. Despite the weaker correlations, A-CTH shows smaller errors than A-TC. In particular, regardless of the baseline, 76% of A-TC CTH estimates fall below the 0.3 km error threshold, while this percentage increases to 79% for A-CTH. Furthermore, except A-TC BD (which is slightly higher), the median RMSEs for ATLID products meet the EarthCARE CTH accuracy requirements of 0.30 km.

The findings of this study indicate an overall improvement in the A-TC and A-CTH products for baseline BD compared to the previous baseline BA. For the A-TC classification product, baseline BD reduces overfilling in inhomogeneous regions, while the A-CTH baseline BD demonstrates increased correlation and reduced errors relative to the reference. Planned improvements in the A-PRO processor that will impact the A-TC product are expected to further decrease overfilling in inhomogeneous regions and may enable more accurate comparisons than those presented in this study. Further, it may also be possible to improve the accuracy of the cloud boundary derived from A-TC by filtering the A-TC results based on the associated retrieved extinction values as overfilled areas tend to be associated with lower and more uncertain retrieved extinctions.

Data availability

The EarthCARE Level 2A baseline BA and BD data is available at ESA's MAAP (Multi-mission Algorithm and Analysis Platform, https://portal.maap.eo.esa.int/, last access: September 9, 2026). GLOVE CPL data is available in Zenodo at https://doi.org/10.5281/zenodo.16807221 (Yorks et al., 2025) .

Code availability

The code used to produce the figures and statistical analyses of this study is publicly available on GitHub (https://github.com/NataliaRoldanHenao/EarthCARE-validation).

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

Author contributions

NRH conceptualized the study, reproduced the experiments, and wrote the initial manuscript. JY conceptualized the study and the experiments. JF contributed to coding and accessing the dataset. JY, EN, AT, and DD helped interpret the data and final results. All authors contributed to writing and editing the manuscript.

Competing interests

At least one of the (co-)authors is a member of the editorial board of Atmospheric Measurement Techniques.

Acknowledgements

We sincerely acknowledge all individuals who contributed to the success of GLOVE. We extend special thanks to CPL team members, the University of Iowa staff and students, and all NASA ER-2 pilots, navigators, staff, forecasters, and crew members. We also wish to acknowledge everyone who contributed to the EarthCARE mission. We acknowledge the use of Grammarly (https://www.grammarly.com/) to assist with grammar correction and improve the academic tone.

Financial support

This study was supported by the Division of Earth Sciences. GLOVE was funded by the ICESat-2 mission (Thorsten Markus, Tom Neumann, and Nathan Kurtz), NASA's Earth Science Research Program (Jack Kaye), NASA's Weather and Atmospheric Dynamics Research Program (Will McCarty), NASA's Radiation Sciences Program (Hal Maring). This research was supported by the NASA Grant 80NSSC23M0011 awarded to the University of Maryland.

References

Donovan, D. P., van Zadelhoff, G. J., and Wang, P.: The EarthCARE lidar cloud and aerosol profile processor (A-PRO): the A-AER, A-EBD, A-TC, and A-ICE products, Atmos. Meas. Tech., 17, https://doi.org/10.5194/amt-17-5301-2024, 2024.

Eisinger, M., Marnas, F., Wallace, K., Kubota, T., Tomiyama, N., Ohno, Y., Tanaka, T., Tomita, E., Wehr, T., and Bernaerts, D.: The EarthCARE mission: science data processing chain overview, Atmos. Meas. Tech., 17, 839–862, https://doi.org/10.5194/amt-17-839-2024, 2024.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

Ellis, T. A., Myers, J., Grant, P., Platnick, S., Guerin, D. C., Fisher, J., Song, K., Kimchi, J., Kilmer, L., LaPorte, D. D., and Moeller, C. C.: The NASA enhanced MODIS airborne simulator, in: Earth Observing Systems XVI, https://doi.org/10.1117/12.894482, 2011.

European Space Agency: EarthCARE Product Data Handbook, https://earthcarehandbook.earth.esa.int, last access: 16 August 2026., 2026.

Illingworth, A. J., Barker, H. W., Beljaars, A., Ceccaldi, M., Chepfer, H., Clerbaux, N., Cole, J., Delanoë, J., Domenech, C., Donovan, D. P., Fukuda, S., Hirakata, M., Hogan, R. J., Huenerbein, A., Kollias, P., Kubota, T., Nakajima, T., Nakajima, T. Y., Nishizawa, T., Ohno, Y., Okamoto, H., Oki, R., Sato, K., Satoh, M., Shephard, M. W., Velázquez-Blázquez, A., Wandinger, U., Wehr, T., Zadelhoff, G.-J. van, Illingworth, A. J., Barker, H. W., Beljaars, A., Ceccaldi, M., Chepfer, H., Clerbaux, N., Cole, J., Delanoë, J., Domenech, C., Donovan, D. P., Fukuda, S., Hirakata, M., Hogan, R. J., Huenerbein, A., Kollias, P., Kubota, T., Nakajima, T., Nakajima, T. Y., Nishizawa, T., Ohno, Y., Okamoto, H., Oki, R., Sato, K., Satoh, M., Shephard, M. W., Velázquez-Blázquez, A., Wandinger, U., Wehr, T., and Zadelhoff, G.-J. van: The EarthCARE Satellite: The Next Step Forward in Global Measurements of Clouds, Aerosols, Precipitation, and Radiation, Bull. Am. Meteorol. Soc., 96, 1311–1332, https://doi.org/10.1175/BAMS-D-12-00227.1, 2015.

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, https://doi.org/10.5194/amt-16-2795-2023, 2023.

Kubota, T., Kikuchi, M., Muto, M., Hashimoto, M., Imura, Y., Maruyama, K., Hoffmann, A., Hummel, T., Koopman, R., Malina, E., Rusli, S., Tzallas, V., von Bismarck, J., and Frommknecht, B.: Summary of the EarthCARE science and validation workshop 2025 -Early achievements and future perspectives for the Earth Cloud Aerosol and Radiation Explorer (EarthCARE) satellite mission-, Journal of the European Meteorological Society, 4, 100037, https://doi.org/10.1016/J.JEMETS.2026.100037, 2026.

Li, L., Heymsfield, G. M., Racette, P. E., Tian, L., and Zenker, E.: A 94-GHz cloud radar system on a NASA high-altitude ER-2 aircraft, J. Atmos. Ocean. Technol., 21, https://doi.org/10.1175/1520-0426(2004)021<1378:AGCRSO>2.0.CO;2, 2004.

McGill, M., Hlavka, D., Hart, W., Scott, V. S., Spinhirne, J., and Schmid, B.: Cloud Physics Lidar: instrument description and initial measurement results, Appl. Opt., 41, 3725–3734, https://doi.org/10.1364/AO.41.003725, 2002.

McGill, M. J., Hlavka, D. L., Hart, W. D., Welton, E. J., and Campbell, J. R.: Airborne lidar measuremens of aerosol optical properties during SAFARI-2000, Journal of Geophysical Research D: Atmospheres, 108, 8493, https://doi.org/https://doi.org/10.1029/2002JD002370, 2003.

McGill, M. J., Vaughan, M. A., Trepte, C. R., Hart, W. D., Hlavka, D. L., Winker, D. M., and Kuehn, R.: Airborne validation of spatial properties measured by the CALIPSO lidar, Journal of Geophysical Research Atmospheres, 112, https://doi.org/10.1029/2007JD008768, 2007.

https://doi.org/10.5194/egusphere-2026-5919
Preprint. Discussion started: 4 November 2026
© Author(s) 2026. CC BY 4.0 License.

de Villele, G., Helière, A., Tzeremes, G., Marnas, F., Wallace, K., Haas, C., Corselle, B., Schillinger, M., Belhadj, T., Schaube, A., Evin, R., Bravetti, P., Da Silva, T. P., Olivier, F., Noyes, C., Chassat, F., De Sanctis, V., and Sapia, A.: ATLID instrument: new UV high-spectral-resolution LiDAR in orbit, https://doi.org/10.1117/12.3072769, 2025.

Walker Mclinden, M. L., Li, L., Heymsfield, G. M., Coon, M., and Emory, A.: The nasa GSFC 94-GHz airborne solid-state cloud radar system (CRS), J. Atmos. Ocean. Technol., 38, https://doi.org/10.1175/JTECH-D-20-0127.1, 2021.

Wandinger, U., Haarig, M., Baars, H., Donovan, D., and Van Zadelhoff, G. J.: Cloud top heights and aerosol layer properties from EarthCARE lidar observations: The A-CTH and A-ALD products, Atmos. Meas. Tech., 16, https://doi.org/10.5194/amt-16-4031-2023, 2023.

Wehr, T., Kubota, T., Tzeremes, G., Wallace, K., Nakatsuka, H., Ohno, Y., Koopman, R., Rusli, S., Kikuchi, M., Eisinger, M., Tanaka, T., Taga, M., Deghaye, P., Tomita, E., and Bernaerts, D.: The EarthCARE mission – science and system overview, Atmos. Meas. Tech., 16, 3581–3608, https://doi.org/10.5194/amt-16-3581-2023, 2023.

Wehr, T. (Ed.): EarthCARE Mission Requirements Document, https://doi.org/10.5270/esa.earthcare-mrd.2006 (last access: 15 July 2026), 2006.

Yorks, J., Selmer, P., Moore, J., Christian, K., Midzak, N., Finlon, J., Cresanti, A., Kuang, S., Finneman, G., Begolka, J., McGill, M., and Nowottnick, E.: Cloud Physics Lidar Level 1 & 2 Data Products from the GLOVE Field Campaign (Version Version 1) [Dataset], https://doi.org/https://doi.org/10.5281/zenodo.16807221, 2025a., 2025.

Yorks, J. E., Hlavka, D. L., Vaughan, M. A., McGill, M. J., Hart, W. D., Rodier, S., and Kuehn, R.: Airborne validation of cirrus cloud properties derived from CALIPSO lidar measurements: Spatial properties, Journal of Geophysical Research: Atmospheres, 116, https://doi.org/10.1029/2011JD015942, 2011a.

Yorks, J. E., Hlavka, D. L., Hart, W. D., McGill, M. J., Yorks, J. E., Hlavka, D. L., Hart, W. D., and McGill, M. J.: Statistics of Cloud Optical Properties from Airborne Lidar Measurements, J. Atmos. Ocean. Technol., 28, 869–883, https://doi.org/10.1175/2011JTECHA1507.1, 2011b.

Yorks, J. E., Selmer, P. A., Kupchock, A., Nowottnick, E. P., Christian, K. E., Rusinek, D., Dacic, N., and McGill, M. J.: Aerosol and Cloud Detection Using Machine Learning Algorithms and Space-Based Lidar Data, Atmosphere 2021, Vol. 12, Page 606, 12, 606, https://doi.org/10.3390/ATMOS12050606, 2021.

Yorks, J. E., Nowottnick, E. P., Platnick, S., Meyer, K. G., Walker McLinden, M., Kacenelenbogen, M. S. F., Christian, K. E., Finlon, J. A., Midzak, N. A., Roldán-Henao, N., Selmer, P. A., McGill, M. J., Dolinar, E. K., Helms, C. N., Koopman, R., von Bismark, J., and Piñol Solé, M.: The GSFC Lidar Observation and Validation Experiment (GLOVE) field campaign, Earth Syst. Sci. Data, 18, 3833–3852, https://doi.org/10.5194/ESSD-18-3833-2026, 2026.

Van Zadelhoff, G. J., Donovan, D. P., and Wang, P.: Detection of aerosol and cloud features for the EarthCARE atmospheric lidar (ATLID): the ATLID FeatureMask (A-FM) product, Atmos. Meas. Tech., 16, 3631–3651, https://doi.org/10.5194/AMT-16-3631-2023, 2023.