MPhys Report

AO06: Validation for Cloud Extinction Measurement of ATLID with Inversion Results from Airborne Lidar

Name: Yixuan Pu,
Supervisors: Dr. Rui Song, and Prof. Roy G. Grainger

(Dated: 28 April 2025)

The FAAM BAe-146-301 atmospheric research aircraft, equipped with a Leosphere ALS450 lidar system, is used in this study to validate cloud optical property data products generated by the Atmospheric Lidar (ATLID) onboard ESA's EarthCARE satellite. A rigorous data-cleaning and calibration procedure is applied to ensure the accuracy of cloud extinction profiles retrieved from the FAAM measurements. The methodology is further cross-validated using a historical flight dataset from 2015, provided by Marenco et al., which shows strong agreement with previously published results. Overall, the FAAM flight datasets support the reliability of ATLID's cloud extinction products. However, minor discrepancies were identified in ATLID's detection of closely spaced cloud layers, highlighting potential limitations in its layer detection algorithm.

I. INTRODUCTION

Light Detection and Ranging (Lidar) is a fundamental remote sensing technique in atmospheric science, providing high-resolution vertical profiles of the optical properties of aerosols, clouds, and trace gases. These measurements are critical for retrieving microphysical properties, evaluating aerosol–cloud interactions, quantifying radiative impacts, and assessing associated climate feedbacks. As such, Lidar data play an essential role in supporting and validating weather and climate models.

A key limitation of conventional elastic backscatter lidar systems is their inability to independently separate particulate (Mie) scattering from aerosols and clouds, and molecular (Rayleigh) scattering from atmospheric gases. This separation is critical for accurately retrieving the optical and microphysical properties of aerosols and clouds. Polarization-based lidars, widely used over the past decade, provide qualitative information on aerosol type and phase through depolarization measurements but do not enable independent retrievals of Mie and Rayleigh components. As a result, cloud and aerosol property retrievals have relied heavily on inversion algorithms[1,2]. Despite methodological improvements, these algorithms remain under-constrained and are highly sensitive to assumptions about the atmospheric state.

To overcome these limitations, High Spectral Resolution Lidar (HSRL) was developed to enable the direct and independent measurement of Mie and Rayleigh scattering. HSRL systems use spectrally selective optical elements – such as a Fabry-Perot interferometer[3] or an iodine vapor absorption filter[4,5] – to isolate the Doppler-broadened Rayleigh signal from the narrowband Mie signal. This spectral discrimination allows for unambiguous retrieval of each scattering component, significantly improving the accuracy and reliability of aerosol and cloud measurements.

Sustained research efforts have advanced the application of HSRL in atmospheric remote sensing[6–9]. A major milestone was the 2024 launch of the European Space Agency's EarthCARE mission, which carries ATLID - the third spaceborne HSRL dedicated to cloud and aerosol profiling[10]. By independently measuring Mie scattering, ATLID enables direct observation of key aerosol and cloud optical properties, such as backscatter and extinction coefficients, and the lidar ratio. This represents a significant advancement over its predecessor, CALIPSO, which relied on elastic backscatter and inversion techniques[11]. ATLID's direct retrieval capability substantially improves the accuracy, reliability, and vertical resolution of aerosol and cloud measurements from space[12], enhancing our ability to study aerosol–cloud interactions and their impact on Earth's radiation budget.

Despite ATLID's advanced capabilities, its operation as a newly developed spaceborne HSRL introduces technical uncertainties that require systematic validation. In particular, a growing body of evidence – from ATLID simulations[13–15] and initial validation results from China's Daqi-1 HSRL mission launched in 2022[16] – suggests that overlying cloud layers may bias the retrieval of optical properties in underlying cloud structures. Consequently, validation campaigns targeting ATLID's performance in multi-layer cloud conditions are essential to ensure the accuracy and reliability of its data products.

As part of the global validation campaign for ATLID, this project will examine several underpasses of the EarthCARE satellite with the FAAM BAe-146-301 atmospheric research aircraft[17] operated by the UK Met Office and the Natural Environment Research Council. The aircraft carries a Leosphere ALS450 polarization-based lidar onboard to perform parallel data collection with the satellite.

Although acquired using traditional lidar systems, the aircraft-based observations offer distinct advantages that make them a valuable benchmark for validating ATLID measurements. In particular, the aircraft's ability to operate beneath upper cloud layers enables targeted, high-resolution sampling of lower cloud structures, facilitating focused assessment of ATLID's performance in complex multi-layer cloud conditions. Integrated with results from concurrent validation campaigns worldwide, this effort contributes to a robust evaluation of ATLID's data quality and supports broader confidence in the consistency and accuracy of its global data products.

II. METHODOLOGY

A. Theoretical Background

This section presents the theoretical background underlying the retrieval algorithms. The goal of the retrieval process is to isolate and quantify the particulate contribution. In this report, variables related to molecular and particulate components are denoted by the subscripts 'm' and 'a', respectively, while quantities without a subscript represent the combined effects of both.

1. The Lidar Equation

The lidar equation forms the foundation of atmospheric lidar retrievals, as it quantitatively describes the molecular and particulate contributions to the received signal. Building on the theoretical framework by Klett (1981)[1], Fernald (1984)[2] formulated an expression for the total backscattered signal a lidar receives from a given slant range z, accounting for optical contributions from both molecules and particles, as:

$$p(z) = K \cdot z^{-2} \left[\beta_m(z) + \beta_a(z)\right] T_m(z)^2 T_a(z)^2 \quad (1)$$

where:

- the backscatter coefficients $\beta_m(z)$ and $\beta_a(z)$ respectively describe the molecular and particulate light-backscattering probability at slant range z;
- transmission functions $T_m(z) = \exp[-\int_0^z \alpha_m(z')\mathrm{d}z']$ and $T_a(z) = \exp[-\int_0^z \alpha_a(z')\mathrm{d}z']$ respectively accounts for the backscattered signal attenuation due to molecules and particles from slant range z to the sensor;
- the extinction coefficients $\alpha_m(z)$ and $\alpha_a(z)$ describe the molecular and particulate signal attenuation probability at slant range z;
- and the lidar constant K represent the overall proportionality constant between the backscattering probability density function and the actual signal count received at the lidar $p(z)$. This factor accounts for both output laser energy and the instrumental losses.

2. The Klett-Fernald Algorithm

The Klett–Fernald algorithm retrieves particle optical properties from the total backscatter signal by numerically solving the lidar equation (Equation 1) under a set of assumptions and boundary conditions. The core assumption of the algorithm is an empirical proportionality between the extinction and backscatter coefficients, expressed as:

$$\alpha_a(z) = S_a \cdot \beta_a(z), \quad \alpha_m(z) = S_m \cdot \beta_m(z); \quad (2)$$

where S_a and S_m are respectively defined as the lidar ratio for particles and molecules.[2]

From this assumption, the lidar equation can be rearranged and integrated from a reference level z_c to solve for the particle extinction coefficients:

$$\alpha_a(z) + \frac{S_a}{S_m}\alpha_m(z) = \frac{x(z) \cdot \exp[-2(S_a - S_m)\int_{z_c}^{z} \beta_m(z')\mathrm{d}z']}{K(z_c) - 2S_a \int_{z_c}^{z} \exp[-2(S_a - S_m)\int_{z_c}^{z} \beta_m(z'')\mathrm{d}z'']\mathrm{d}z'} \quad (3)$$

where $x(z) = z^2 \cdot p(z)$ is the range corrected signal measured by the lidar, and z' and z'' are dummy variables representing the slant range z within the integral expressions. The reference level lidar constant $K(z_c)$ can then be substituted by:

$$K(z_c) = \frac{x(z_c)}{\alpha_a(z_c) + \frac{S_a}{S_m}\alpha_m(z_c)}. \quad (4)$$

The Klett–Fernald algorithm applies a stepwise numerical integration scheme to Equation 3, computing

molecular and particulate optical properties at each step by propagating forward $(I+1)$ or backward $(I-1)$ from a reference point (I):

$$\alpha_a(I+1)+\frac{S_a}{S_m}\alpha_m(I+1)= \frac{x(I+1)\exp[-\psi(I,I+1)]}{K(I)-S_a\{x(I)+x(I+1)\exp[-\psi(I,I+1)]\}\cdot\Delta z} \tag{5}$$

$$\alpha_a(I-1)+\frac{S_a}{S_m}\alpha_m(I-1)= \frac{x(I-1)\exp[\psi(I-1,I)]}{K(I)+S_a\{x(I)+x(I-1)\exp[-\psi(I-1,I)]\}\cdot\Delta z} \tag{6}$$

where $\psi(I,I+1)=(S_a-S_m)\,[\beta_m(I)+\beta_m(I+1)]$ accounts for the exponential terms in Equation 3.

The Klett–Fernald algorithm is a widely adopted and effective method for retrieving particle optical properties from lidar signals. However, due to its stepwise integration approach starting from a defined boundary condition, the algorithm is highly sensitive to instrumental noise in between integration steps, as well as to uncertainties or errors associated with the initial boundary condition. This necessitates meticulous noise filtering and rigorous instrument calibration to ensure the reliability and accuracy of retrieved optical properties. The procedures implemented in this study to achieve these are detailed in the following section.

Like many inverse retrieval problems, determining the particle extinction coefficient from lidar measurements using the Klett-Fernald algorithm is inherently ill-posed. This arises from the fact that the lidar equation includes two unknowns – extinction and backscatter coefficients – within a single measured quantity. As a result, the retrieval requires an assumed lidar ratio, which is typically uncertain or unknown, leading to infinitely many possible solutions depending on the chosen value.

To address this, the present study employs lidar ratio values derived from ATLID's HSRL measurements to improve retrieval accuracy.

3. Signal Depolarization

While the Klett–Fernald formulation treats the backscattered signal as a scalar quantity, modern polarization-sensitive lidars, such as the Leosphere ALS450 used in this project, study the backscattered signal in orthogonal polarization components relative to its transmitted laser beam, $\mathbf{p}_{\text{nat}}=(p_{\parallel},\,p_{\perp})$, where:

- **the co-polarized component** $(p_{\parallel})$ is aligned with the transmitted beam's polarization plane, primarily resulting from non-depolarizing scatterers like atmospheric molecules and spherical particles;
- **the cross-polarized component** $(p_{\perp})$ is orthogonal to the transmitted beam's polarization plane, arising from depolarizing scatterers such as non-spherical particles.

The total backscattered signal is the sum of these components:

$$p=p_{\parallel}+p_{\perp}. \tag{7}$$

The backscattered signal component in each polarization direction satisfies the lidar equation independently. In particular, the corresponding total backscatter coefficients in each polarization direction – $\beta_{\parallel}(z)$ and $\beta_{\perp}(z)$ – quantify the likelihood of detecting co- and cross-polarized returns from a range z. Their ratio defines the linear volume depolarization ratio:

$$\delta^V=\frac{\beta_{\perp}}{\beta_{\parallel}}=\frac{p_{\perp}}{p_{\parallel}}, \tag{8}$$

which characterizes the depolarization behavior of atmospheric scatterers and serves as a calibration reference.

Throughout this report, subscripts '$\parallel$' and '$\perp$' consistently denote quantities in the naturally defined co- and cross-polarized frames, relative to the transmitted laser beam's polarization axis.

B. Instrumental Specifications

This subsection outlines key specifications of the FAAM BAe-146-301 aircraft and Leosphere ALS450 lidar.

The Leosphere lidar measures atmospheric backscatter on the FAAM aircraft in nadir (downward) viewing geometry[19]. Technical specifications for the lidar system are summarized in Table I, with further relevant details discussed in the sub-subsections that follow.

1. Depolarization Channels

Depolarization separation in the Leosphere ALS450 is achieved using a polarization beam splitter (PBS) positioned after the lidar receiver. The instrument reads the backscattered signal from two orthogonal output polarization channels from the PBS, $\mathbf{p}_{\text{out}}=(p_T,\,p_R)$, where:

TABLE I. Technical information for Leosphere ALS450 lidar.

Parameter	Specification[a]
Transmitter	
Laser Type	Nd:YAG
Transmitted Wavelength	354.7 nm
Pulse Repetition Frequency	20 Hz
Pulse Energy	16 mJ (typical)
Pulse Duration	6 ns
Transmitted Beam Diameter	0.025 m
Transmitted Beam Geometry	Biaxial to receiver (manual steering)
Receiver	
Lens Diameter	0.15 m
Height Resolution	15 m
Raw Height Resolution	1.5 m
Polarization Beam Splitter (PBS)[b]	
Parallel Transmittance (T_v)	1.0
Perpendicular Transmittance (T_h)	0.00148 (14815)
Parallel Reflectance (R_v)	0.02475 (1952)
Perpendicular Reflectance (R_h)	1.0

[a] The values and comments are documented by the Chilbolton Group of Rutherford Appleton Laboratory[18].
[b] The PBS parameters are provided by Stuart Fox.

- **the transmitted channel** (p_T) captures the component of the signal transmitted through the PBS;
- **the reflected channel** (p_R) captures the component reflected by the PBS.

In ideal situation, the channel readings $\mathbf{p}_{\text{out}}$ are equivalent to the naturally defined polarization components $\mathbf{p}_{\text{nat}}$, such that:

$$p = p_{\parallel} + p_{\perp} = p_T + p_R \tag{9}$$

However, in practice, PBS - transmitted beam misalignments and non-ideal PBS performance will introduce a relative amplification factor V^* between the detection channels. As a result, the total backscattered signal is expressed as:

$$p = p_{\parallel} + p_{\perp} = V^* p_T + p_R \tag{10}$$

To account for this effect, a polarization calibration method based on the theoretical framework developed by Freudenthaler et al. (2009)[20] is implemented. The methodology details are presented in Section II.C.3.

2. Data Structure

Before describing the specific processing methods used in this study, it is useful to outline the structure of the raw data collected by the FAAM aircraft. This analysis includes three datasets from the 2024 validation campaign, along with one legacy dataset from the same aircraft, provided by Start Fox. A summary of the datasets is presented in Table II.

These datasets consist of vertical backscatter profiles recorded by the onboard lidar system. Each profile captures returns over a fixed range that extends below the aircraft's altitude. As a result, the raw profiles include apparent signals from below the surface, which do not correspond to atmospheric returns but represent background instrumental noise. These 'below-ground' signals are a known characteristic of the raw data and are corrected during post-processing. An example is provided in the Appendix.

C. Calibration Methodology

This subsection outlines the calibration methodology employed in this study, comprising the following steps.

1. **Data Preprocessing**: Prepares the raw signal data for calibration.
2. **Background Error Calibration**: Corrects for environment-dependent background systematic errors in each polarization channel.
3. **Polarization Amplification Factor Calibration**: Accounts for the cross-channel gain imbalance and optical misalignment in the polarization detection system.
4. **Lidar Constant Calibration**: Determines the proportionality constant relating the measured signal to the scattering probability density function.

These steps are summarized in the flowchart shown in Figure 1 and are described in detail in the following sub-subsections.

1. Data Preprocessing

To optimize accuracy, each flight dataset is preprocessed in two parallel copies, tailored respectively for calibration and retrieval.

The calibration dataset is restricted to a narrow altitude band in the upper troposphere to avoid lower-level aerosols and ensure accurate characterization of clear-sky Rayleigh scattering. Instrumental noise is removed

TABLE II. Summary of FAAM flight datasets analyzed in this study.

Flight ID[a]	Date	Latitude Range[b]		Longitude Range[b]		Height Grid[c]	Time Grid[c]
		Min	*Max*	*Min*	*Max*		
B895	2015-03-13	55.719° N	60.482° N	−6.154° E	−2.614° E	45 m	2.5 s
C394	2024-11-05	52.910° N	60.929° N	−9.301° E	−1.326° E	15 m	2.5 s
C395	2024-11-13	52.107° N	57.621° N	−1.835° E	0.026° E	15 m	2.5 s
C397	2024-11-22	52.159° N	57.791° N	−1.394° E	0.731° E	15 m	2.5 s

[a] Dataset B895 was provided by Stuart Fox; other datasets were collected during the 2024 validation campaign.
[b] The measurement region boundaries are defined by the minimum and maximum latitude and longitude observed during each flight.
[c] The height and time grids represent the vertical and temporal resolutions, respectively.

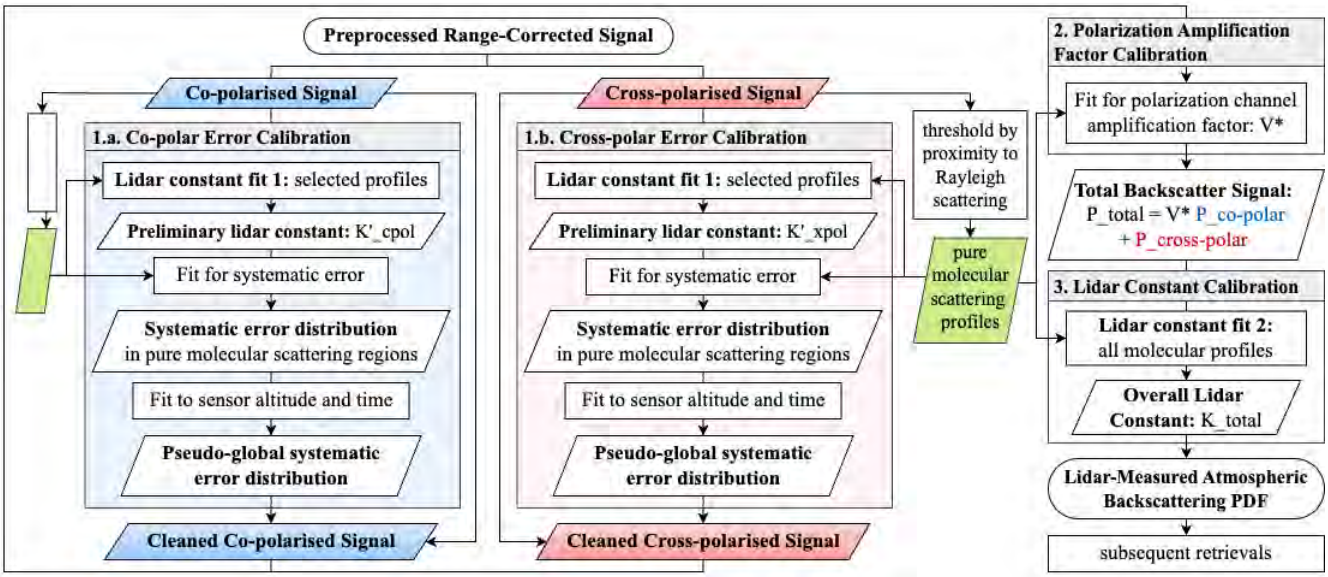

FIG. 1. Flowchart outlining the calibration procedure applied to the co-polarized and cross-polarized datasets. Each channel first undergoes initial calibration to derive a pseudo-global background error profile, which is then applied to enable two subsequent calibration steps. See Section II.C for details.

by subtracting a moving average of the below-ground signal introduced in Section II.B.2. No additional filtering is applied to preserve the integrity of the calibration.

The retrieval dataset, in contrast, spans the full available altitude range to capture cloud structures. After the same noise removal step, a low-pass Fourier filter is applied to improve stability for the Klett–Fernald inversion. Specifically, an elliptical Gaussian mask is used in the Fourier domain to suppress high-frequency noise along both temporal and vertical dimensions, while preserving relevant cloud features.

2. *Background Systematic Error Calibration*

This sub-subsection describes the background systematic error calibration procedure, performed independently for the co-polarized (transmitted) and cross-polarized (reflected) channels. The overall workflow is summarized in Figure 1, and consists of the following steps.

(a) Compute the Rayleigh scattering PDF in the lidar's range frame for the full flight.

(b) Identify and extract clear-sky profiles dominated by Rayleigh scattering, then interpolate them to construct a mapping between signal strength and Rayleigh PDF.

(c) Estimate an initial lidar constant (K') from stable Rayleigh profiles, then model systematic errors across all profiles as a function of aircraft altitude and time to derive a pseudo-global background error profile.

A detailed explanation of each step is provided in the

following subsections of this sub-subsection.

(a) Rayleigh Signal Modeling

In pure molecular scattering conditions, the lidar equation (Equation 1) introduced in Section II.A.1 simplifies to:

$$\begin{aligned} z^2 p(z) &= K \cdot \beta_m(z) T_m(z)^2 \\ &= K \cdot \beta_m(z) \exp\left[-2\int_0^z \alpha_m(z')\mathrm{d}z'\right]. \end{aligned} \tag{11}$$

Here, the molecular scattering lidar ratio is defined as $S_m = \alpha_m(z)/\beta_m(z) = 8\pi/3$, derived from the Rayleigh scattering phase function. The molecular extinction coefficient at a given atmospheric height h is computed as:

$$\alpha_m(h) = \rho(h) \cdot \sigma_e \tag{12}$$

where $\rho(h)$ is the atmospheric molecular density, and σ_e is the effective elastic scattering cross-section at the lidar's operating wavelength, calculated using the empirical formulation of Bucholtz (1995)[21].

For airborne lidar systems, the physical altitude $h(t)$ corresponding to slant range z at time t is given by:

$$h(t) = A_s(t) - z \tag{13}$$

where $A_s(t)$ denotes the aircraft altitude. To simulate the Rayleigh backscatter profile across altitude-varying segments of the dataset, the height-dependent molecular density profile must first be interpolated from the physical height grid to the slant range grid before applying Equation 11.

(b) Calibration Region Extraction

Once Rayleigh signals are simulated across the dataset, calibration-suitable regions are identified by evaluating each lidar profile's similarity to ideal Rayleigh scattering behaviour using a Gaussian Mixture Model. Consecutive profiles classified as good fits are grouped into continuous calibration regions. Additional details and illustrative plots for this process can be found in Appendix.

Calibrating on a fixed physical height grid is limited by variations in aircraft altitude, which introduce misalignment between profiles and inconsistencies in Rayleigh signal values along the slant range. This constraint significantly reduces the number of usable calibration points – particularly in cloudy conditions where clear-sky references are limited.

To address this, calibration is performed on an evenly spaced Rayleigh signal grid, where the Rayleigh backscatter value serves as the reference coordinate instead of physical height. This signal-based approach eliminates altitude misalignment, enabling consistent use of all valid calibration points and improving the accuracy and stability of the retrieved calibration constant.

(c) Modelling Systematic Error Profile

As introduced earlier and illustrated in Figure 1, an initial calibration step is performed for each depolarization channel to account for systematic errors that vary with sensor altitude and time. A linear regression model with a non-zero intercept is fitted to a subset of stable calibration profiles that closely match the expected Rayleigh scattering signal. This regression provides an initial estimate of the channel-specific lidar constant (K'), which serves as a provisional scaling factor for the backscatter signal.

Using this trial constant, linear fits are performed across all calibration profiles, and the resulting y-intercepts representing deviations from the Rayleigh model are interpreted as empirical estimates of systematic error. These intercepts are then modeled as a continuous function of sensor altitude $A_s(t)$ and time t, yielding a pseudo-global error profile for each channel:

$$\Delta(A_s(t), t) = a \cdot A_s(t) + b + c \cdot t + d \cdot t^2 + e \cdot t^3. \tag{14}$$

Here, $\Delta(A_s(t), t)$ denotes the estimated systematic error, and $A_s(t)$ is the aircraft's altitude. The modelled error profiles are subsequently subtracted from the corresponding depolarization channel datasets prior to the next calibration stage.

3. Polarization Amplification Factor Calibration

With the pseudo-global correction for systematic errors applied to each depolarization channel, the signal measurements are sufficiently calibrated to determine the relative polarization amplification factor V^*, as defined in Equation 10.

Figure 2 illustrates the geometric relationship between polarization components of the backscattered beam when an angular misalignment ϕ exists between the polarization plane of the transmitted laser beam and the incident plane of the polarization beam split-

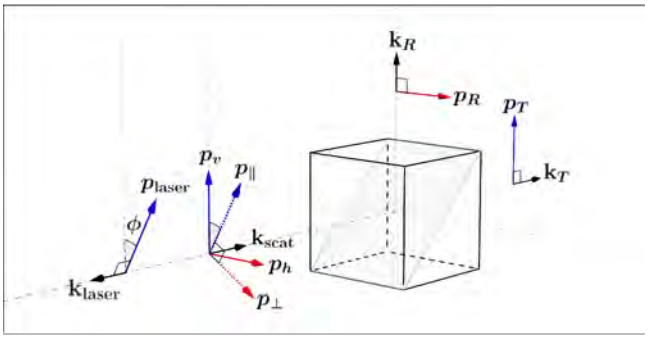

FIG. 2. Schematic diagram of the geometric relationship of measured signals in beam splitter polarization channels (p_T, p_R) to the real co-polarized and cross-polarized signals $(p_\parallel, p_\perp)$.

ter (PBS), following the framework of Freudenthaler et al. (2009)[20]. In this configuration, the polarization state of the backscattered beam is resolved relative to two reference planes:

- **Laser polarization plane**: defines the natural or physically meaningful co-polarized and cross-polarized components, $\mathbf{p}_{\text{nat}} = (p_\parallel, p_\perp)$.
- **PBS incident plane**: determines the parallel and perpendicular components entering the PBS, $\mathbf{p}_{\text{inc}} = (p_v, p_h)$.

Using standard trigonometric projection, the transformation between these two frames is:

$$\begin{bmatrix} p_v(\phi) \\ p_h(\phi) \end{bmatrix} = \begin{bmatrix} \sin^2(\phi) & \cos^2(\phi) \\ \cos^2(\phi) & \sin^2(\phi) \end{bmatrix} \begin{bmatrix} p_\parallel \\ p_\perp \end{bmatrix} \tag{15}$$

For a PBS with reflectivities R_v and R_h, and transmittances T_v and T_h for polarization components p_v and p_h, the measured signals in the reflected and transmitted channels are:

$$\begin{bmatrix} p_T \\ p_R \end{bmatrix} = \begin{bmatrix} T_v V_T & T_h V_T \\ R_v V_R & R_h V_R \end{bmatrix} \begin{bmatrix} p_v(\phi) \\ p_h(\phi) \end{bmatrix} \tag{16}$$

where V_T and V_R represent the combined optical and electronic amplification factors for the transmitted and reflected channels, respectively. The relative depolarization amplification factor is defined as:

$$V^* = \frac{V_R}{V_T},$$

and the measured signal ratio is:

$$\delta^*(\phi) = \frac{p_R(\phi)}{p_T(\phi)}.$$

Using the definition for volume depolarization ratio $\delta^V = p_\perp / p_\parallel$ from Equation 8, and substituting into Equations 15 and 16, the measured signal ratio becomes:

$$\delta^*(\phi) = V^* \frac{[1+\delta^V \tan^2(\phi)]R_v + [\tan^2(\phi)+\delta^V]R_h}{[1+\delta^V \tan^2(\phi)]T_v + [\tan^2(\phi)+\delta^V]T_h}. \tag{17}$$

This expression provides the basis for determining V^* by fitting to observed signal ratios $\delta^*(\phi)$, accounting for angular misalignment and PBS imperfections.

For pure molecular scattering, the volume depolarization ratio is a constant determined by the polarizability anisotropy of air molecules and the probing wavelength. At the operating wavelength of the Leosphere ALS450 lidar used in this study, the molecular volume depolarization ratio is approximately: $\delta_m^V(\lambda = 354.7\text{ nm}) \approx 0.00457$.

As shown in Table I, the lidar's output beam is manually aligned to maintain a biaxial configuration with the receiver axis before each flight. As a result, the angular misalignment φ is expected to be small and close to zero at the start of each flight. Under this assumption, Equation 17 can be expanded for small $\varphi \sim 0.1$ rad as:

$$\delta^*(\varphi(t)) = V^* \frac{[\delta^V R_v + R_h]\varphi(t)^2 + [R_v + \delta^V R_h]}{[\delta^V T_v + T_h]\varphi(t)^2 + [T_v + \delta^V T_h]} + \mathscr{O}(\varphi(t)^3), \tag{18}$$

Assuming that the time dependence of all other parameters is absorbed into the modeled angular misalignment function $\varphi(t)$, and using the boundary condition $\varphi(0) = 0$, the relative depolarization amplification factor V^* is first estimated from the ratio of reflected to transmitted signals in the initial 10 clear-sky, pure-Rayleigh profiles. This value is then used to derive a pseudo-global fit for angular misalignment using the same empirical form as for background error correction:

$$\varphi(A_s(t), t) = a \cdot A_s(t) + b + c \cdot t + d \cdot t^2 + e \cdot t^3, \tag{19}$$

where each profile's angular misalignment $\varphi(t)$ is estimated from the gradient of a direct-proportionality fit between the reflected and transmitted channel signals.

Using the derived angular misalignment distribution $\varphi(t)$ and amplification factor V^*, the true co-polarized and cross-polarized components of the backscattered signal, $p_\parallel(t)$ and $p_\perp(t)$, are recovered by inverting Equations 15 and 16. The total backscattered signal

is then given by:

$$p(t,z) = p_{\parallel}(t,z) + p_{\perp}(t,z), \tag{20}$$

where t is the time of each vertical profile measurement and z is the range from the sensor, as defined in Section II.B.2.

4. Lidar Constant Calibration

With the total backscattered signal fully calibrated, the overall lidar proportionality constant is derived from a direct linear fit in Rayleigh PDF space. Strong agreement across molecular scattering profiles enables a robust global estimate of the lidar constant K, with an illustrative example provided in the Appendix. Applying this constant to the calibrated signal yields accurate estimates of the backscatter PDF beneath the aircraft, forming the basis for cloud and aerosol retrievals using the Klett–Fernald algorithm.

D. Cloud Optical Properties Retrieval

Cloud extinction cross-sections are retrieved from the calibrated backscatter PDF using the Klett–Fernald algorithm, as outlined in Section II.A.2, and applied to the retrieval datasets described in Section II.C.1.

For each flight, a Rayleigh-dominated layer at a reference height H_c beneath the aircraft is selected to serve as the lower boundary condition for the inversion. Cloud optical properties above this boundary are first retrieved using the backward Klett–Fernald algorithm (Equation 6), proceeding in a bottom-up direction. For each profile, the transmission function representing attenuation above the boundary is calculated as:

$$T_{\text{upper}}(t) = \exp\left\{-2\int_0^{z_c(t)} \left[\alpha_a(t,z') + \alpha_m(t,z')\right] \mathrm{d}z'\right\}, \tag{21}$$

where $z_c(t) = A_s(t) - H_c$ is the slant range to the molecular reference layer, and $A_s(t)$ is the sensor altitude at time t.

The measured backscatter PDF from atmospheric layers below the boundary is then corrected using $T_{\text{upper}}(t)$ to account for attenuation by the overlying cloud. Following this correction, particle optical properties below the boundary are retrieved in a top-down direction using the forward Klett–Fernald algorithm (Equation 5).

III. RESULTS AND DISCUSSION

Cloud extinction profiles were retrieved using the methodology described in Section II.D for three key validation flights: C394, C395, and C397, as well as the legacy flight B895, previously analyzed by Marenco et al., and independently validated in other studies. Measurement dates and regions for these flights can be found in Table II.

For validation flights C395 and C397, meaningful quantitative comparisons with satellite observations were not possible due to significant temporal offsets between the airborne and spaceborne measurements. Full cloud extinction profiles for these flights, as well as for flight B895, are provided in Appendix C.

Flight C394 exhibits strong spatial and temporal overlap between the satellite and airborne observations. A persistent altostratus cloud layer is present near 2 km altitude for most of the measurement period, with a continuous cirrus layer above, beginning at a cloud base around 10 km. This configuration provides ideal conditions for evaluating the potential bias introduced by upper-level cloud layers in ATLID measurements, as identified in previous studies[13–16].

Figure 3 presents the comparison between cloud extinction profiles retrieved from FAAM measurements and those observed by ATLID. In the first subplot, detached lower-level marine stratocumulus clouds are evident below the altostratus layer, centered around 1 km altitude. However, this multi-layered structure is largely absent in ATLID's observations, acquired 1.5 to 2 hours later during the EarthCARE overpass. In ATLID's profiles, only a single cloud layer below 3 km is typically detected. While some of this discrepancy may be attributed to boundary layer evolution—such as vertical mixing or entrainment—substantial cloud dissipation or vertical development within this short time frame appears unlikely, given the cold sea surface temperatures and prevailing wind conditions in November.

Vertical profiles within the region of approximate spatial and temporal overlap between EarthCARE and FAAM measurements have been isolated for quantitative comparison. An averaged extinction profile over this region and time window is presented in Figure 4. Good agreement is observed in the upper cloud layer; however, in the lower layer, the FAAM lidar resolves two closely spaced cloud layers, whereas ATLID detects only the lower one. Further examination of ATLID data indicates that when multiple cloud layers are vertically close – such as in the latitude ranges

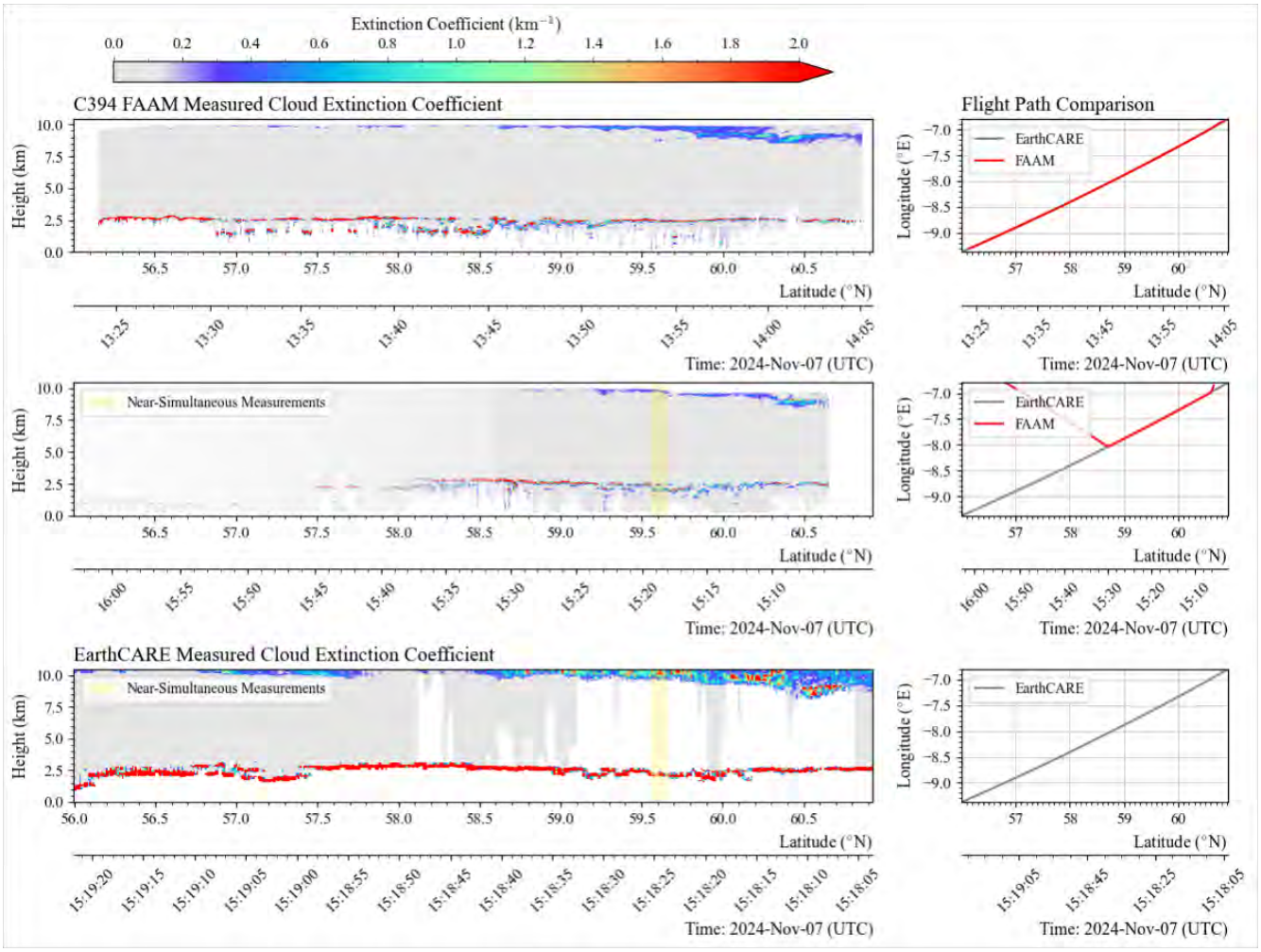

FIG. 3. Comparison of cloud extinction profiles retrieved from FAAM airborne lidar measurements (Flight C394) and those measured by ATLID aboard EarthCARE on 7 November 2024. The FAAM data are segmented based on the aircraft's heading angle, with only segments containing meaningful cloud returns shown. The left-hand subplots display the retrieved cloud extinction coefficient profiles, while the right-hand subplots show the corresponding flight paths for each segment relative to EarthCARE's satellite overpass. Regions of near-complete spatial and temporal overlap between the aircraft and satellite observations are highlighted in yellow. Additional data from segments with partial spatial divergence from the satellite track are included for qualitative comparison; their visual opacity is reduced in proportion to the degree of spatial separation.

57.1°N-57.2°N, 58.0°N-58.2°N, and 59.55°N-59.65°N – ATLID typically resolves either the upper or lower layer, but not both.

This observation may indicate a limitation in ATLID's layer detection algorithm, particularly when resolving closely spaced or vertically overlapping cloud structures. If the algorithm consistently fails to identify multiple adjacent cloud layers – detecting only the optically dominant one – this can introduce significant inaccuracies into its data products. Such misrepresentation affects the derived cloud optical properties, including extinction profiles, cloud-top and cloud-base heights, and layer thicknesses.

These inaccuracies can propagate into higher-level products used in climate modeling, weather forecasting, and cloud–aerosol interaction studies. For instance, underestimating the number of cloud layers or misplacing their vertical positions could bias satellite-derived estimates of radiative forcing, cloud cover, or vertical heating rates. Moreover, these issues may limit the utility of ATLID data in validating atmospheric models or in synergistic retrievals that rely on precise vertical cloud structure (e.g., with EarthCARE's other sensors like CPR or MSI).

Addressing these limitations may require improvements to the layer detection algorithm, such as refining sensitivity thresholds, increasing vertical resolution through post-processing, or incorporating contex-

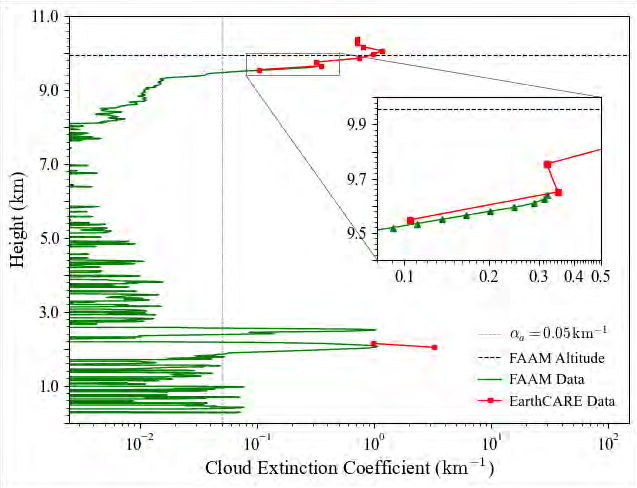

FIG. 4. Comparison of cloud extinction profiles retrieved from FAAM measurements and ATLID observations at 15:18 UTC on 7 November 2024. Profiles are averaged over the spatial domain bounded by (59.63° N, −7.54° E) and (59.67° N, −7.52° E).

tual information from collocated instruments to better distinguish closely packed layers.

IV. CONCLUSION

This study presents a comprehensive methodology for retrieving cloud extinction profiles from FAAM's airborne lidar measurements. Through a carefully calibrated application of the Klett–Fernald algorithm, supported by polarization-channel-specific systematic error correction, depolarization amplification factor correction, and Rayleigh-based lidar constant calibration; accurate high-resolution cloud extinction profiles were retrieved for multiple validation flights.

The retrieved profiles are used as validation for ATLID data. In particular, the dataset from FAAM flight C394 was able to identify ATLID's discrepancies which arise in scenes with closely spaced cloud structures. These inconsistencies, most notably ATLID's occasional omission of one of the cloud layers, suggest potential limitations in its layer detection algorithm – especially under complex vertical cloud distributions. Such limitations may impact the accuracy of ATLID-derived products, thus highlighting the value of high-resolution airborne measurements for satellite validation.

Overall, the results confirm the general reliability of ATLID's data products. However, the minor discrepancies in ATLID's cloud layer detection suggest that continued cross-platform validation is also important to improve the fidelity of spaceborne lidar observations in operational and research applications.

[1] J. D. Klett, "Stable analytical inversion solution for processing lidar returns," Applied Optics **20**, 211 (1981).
[2] F. G. Fernald, "Analysis of atmospheric lidar observations: some comments," Applied Optics **23**, 652 (1984).
[3] S. T. Shipley, D. H. Tracy, E. W. Eloranta, J. T. Trauger, J. T. Sroga, F. L. Roesler, and J. A. Weinman, "High spectral resolution lidar to measure optical scattering properties of atmospheric aerosols. 1: Theory and instrumentation," Appl. Opt. **22**, 3716–3724 (1983).
[4] P. Piironen and E. W. Eloranta, "Demonstration of a high-spectral-resolution lidar based on an iodine absorption filter," Opt. Lett. **19**, 234–236 (1994).
[5] J. W. Hair, L. M. Caldwell, D. A. Krueger, and C.-Y. She, "High-spectral-resolution lidar with iodine-vapor filters: Measurement of atmospheric-state and aerosol profiles," Appl. Opt. **40**, 5280–5294 (2001).
[6] E. Fokitis, P. Fetfatzis, A. Georgakopoulou, S. Maltezos, and A. Aravantinos, "Design of a high spectral resolution lidar for atmospheric monitoring in eas detection experiments," Nuclear Physics B - Proceedings Supplements **190**, 261–265 (2009).
[7] J. Goldsmith, *High Spectral Resolution Lidar Instrument Handbook*, Atmospheric Radiation Measurement (ARM), The U.S. Department of Energy (2016).
[8] Z. Zang, X. Shen, Z. Zheng, Y. Zhang, Y. Zhou, N. Wang, L. Wu, and D. Liu, "Design of a high-spectral-resolution lidar for atmospheric temperature measurement down to the near ground," Appl. Opt. **58**, 9651–9661 (2019).
[9] X. Chen, G. Dai, S. Wu, J. Liu, B. Yin, Q. Wang, Z. Zhang, S. Qin, and X. Wang, "Coherent high-spectral-resolution lidar for the measurement of the atmospheric mie–rayleigh–brillouin backscatter spectrum," Opt. Express **30**, 38060–38076 (2022).
[10] *ATLID design*, European Space Agency (ESA) (2024).
[11] S. A. Young and M. A. Vaughan, "The retrieval of profiles of particulate extinction from cloud-aerosol lidar infrared pathfinder satellite observations (calipso) data: Algorithm description," Journal of Atmospheric and Oceanic Technology **26**, 1105–1119 (2009).
[12] N. Sugimoto, Y. Jin, T. Nishizawa, A. Shimizu, H. Iwai, M. Aoki, K. Yasunaga, K. Yumimoto, H. Okamoto, H. Irie, *et al.*, "Validation of earthcare atlid using the ground-based lidar network ad-net," in *Lidar and Optical Remote Sensing for Environmental Monitoring XVII*, Vol. 13265 (SPIE, 2024) pp. 50–55.
[13] D. Donovan, "The expected impact of multiple scattering on atlid signals," EPJ Web of Conferences **119**, 01006 (2016).
[14] D. P. Donovan, G.-J. van Zadelhoff, and P. Wang, "The earthcare lidar cloud and aerosol profile processor (a-pro): the a-aer, a-ebd, a-tc, and a-ice products," Atmospheric Measurement Techniques **17**, 5301–5340 (2024).
[15] S. L. Mason, H. W. Barker, J. N. S. Cole, N. Docter, D. P. Donovan, R. J. Hogan, A. Hünerbein, P. Kollias, B. Puigdomènech Treserras, Z. Qu, U. Wandinger, and G.-J. van Zadelhoff, "An intercomparison of earthcare cloud, aerosol, and precipitation retrieval products," Atmospheric Measurement Techniques **17**, 875–898 (2024).
[16] Q. Liu, Z. Huang, J. Liu, W. Chen, Q. Dong, S. Wu, G. Dai, M. Li, W. Li, Z. Li, X. Song, and Y. Xie, "Validation of initial observation from the first spaceborne high-spectral-resolution lidar with a ground-based lidar network," Atmospheric Measurement Techniques **17**, 1403–1417 (2024).
[17] Facility for Airborne Atmospheric Measurements (FAAM), "Faam operated bae-146-301 atmospheric research aircraft," (2018), accessed: 2025-04-25.
[18] J. Jeffery, *Technical description of the Chilbolton 'CAMRa' radar*, Chilbolton Group, Rutherford Appleton Laboratory, Science and Technology Facilities Council (Unkown).
[19] Met Office, "Leosphere als450 lidar instrument record," (2018), accessed: 2025-04-25.
[20] V. Freudenthaler, M. Esselborn, M. Wiegner, B. Heese, M. Tesche, A. Ansmann, D. Müller, D. Althausen, M. Wirth, A. Fix, and et al., "Depolarization ratio profiling at several wavelengths in pure saharan dust during samum 2006," Tellus B: Chemical and Physical Meteorology **61**, 165 (2009).
[21] A. Bucholtz, "Rayleigh-scattering calculations for the terrestrial atmosphere," Appl. Opt. **34**, 2765–2773 (1995).

APPENDIX

A. RAW DATA STRUCTURE

This section presents an example raw data plot from the B895 dataset, provided by Stuart Fox, to illustrate the data structure used in this project.

Figure 5 shows the raw cross-polarized, range-corrected backscatter signal recorded by the FAAM aircraft during flight B895, plotted as a function of time and height. The colour scale represents the analogue signal count intensity. The red line traces the aircraft's vertical position throughout the flight.

Below the zero height line, apparent signals from negative altitudes represent background instrumental noise, as the lidar continues to sample fixed range bins beyond the Earth's surface. These 'below-ground' signals are a characteristic feature of raw airborne lidar data as mentioned in Section II.B.2 and are corrected during preprocessing, as described in Section II.C.1.

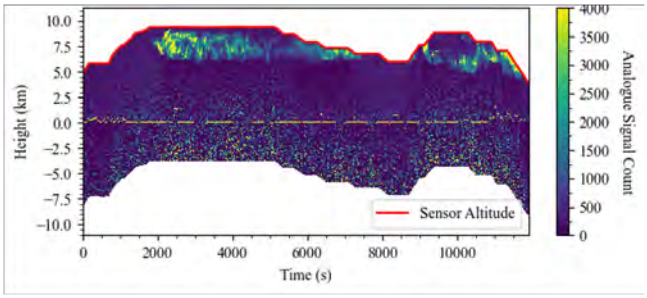

FIG. 5. Raw range-corrected signal dataset B895 by FAAM flight on 13th March 2015, plotted against height above sea level, supplied by Stuart Fox.

B. CALIBRATION

This section presents example figures to aid the reader's understanding of the calibration process.

Figure 6 illustrates the extraction of calibration regions from the processed lidar data. The top panel shows the range-corrected backscatter signal across the full flight, while the middle panel highlights the selected calibration regions where the measured signal closely follows Rayleigh scattering behavior. The bottom panel shows the corresponding Rayleigh scattering profiles used as reference during calibration. The dashed line in each plot traces the aircraft's altitude, demonstrating the altitude dependence of available calibration regions throughout the flight.

This section also presents an example of the background systematic error calibration results. Figure 7 shows a pseudo-global fit of systematic errors for the cross-polarized channel, derived from the four datasets analyzed in this study. The data points represent the estimated systematic errors extracted from individual calibration profiles, while the fitted curves model the variation of background error as a continuous function of sensor altitude and time. These pseudo-global error models are applied during preprocessing to correct the measured backscatter signals prior to further calibration steps.

Figure 8 illustrates the unified data collected from all identified calibration regions across the flight. Each point represents a range-corrected backscatter signal measurement and its corresponding molecular scattering reference value, with different colors indicating different calibration regions. The horizontal error bars reflect the uncertainty in the measured signal within each region. A linear fit to the combined dataset, shown as the fitted calibration model, is used to determine the final lidar calibration constant for the channel. This unified fitting approach improves robustness by leveraging all available clear-sky calibration points across the flight.

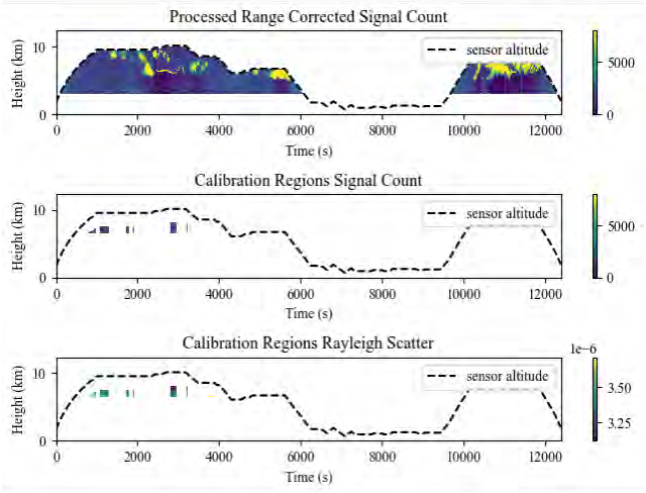

FIG. 6. Example calibration region extraction for one of the validation flights (FAAM-C395). The first subplot shows the preprocessed calibration dataset. The second subplot shows the extracted regions with calibration-suitable conditions. The last subplot shows the simulated Rayleigh scattering signal in the calibration regions. The dotted line marks the altitude of the aircraft.

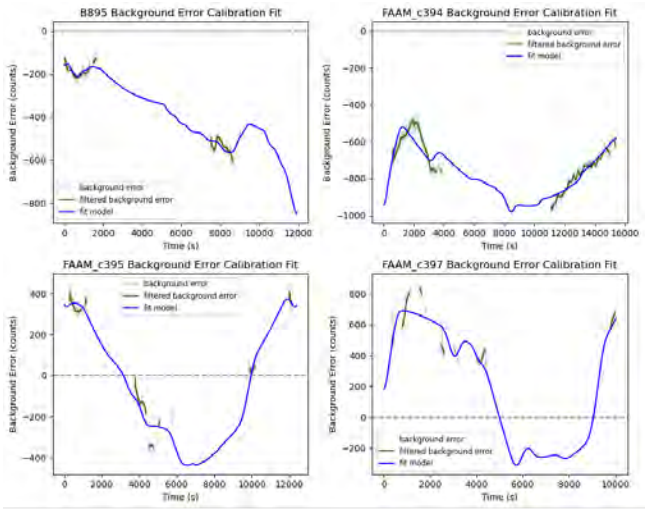

FIG. 7. Background systematic error fit in cross-polar channel for four measurement flights by the FAAM-BAe-146-301 aircraft. Flights C394, C395 and C397 are measurements of the current validation campaign with data collection in November 2024. Flight B895 is included as a reference measurement with data collected in March 2015.

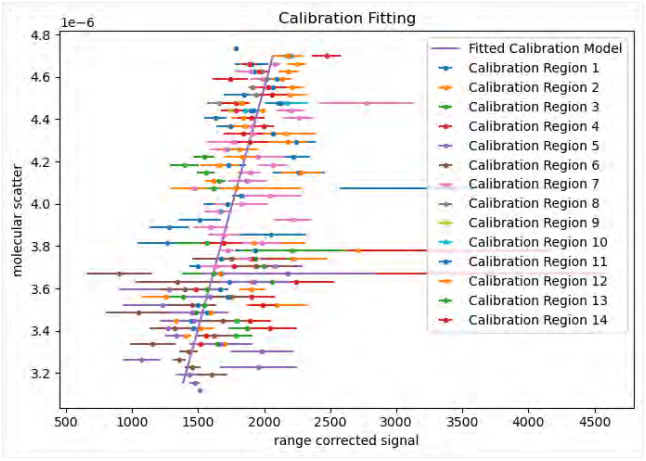

FIG. 8. Final calibration plot for flight FAAM-C395. (See figure 6 for spatial plots of the regions - the regions are labeled in increasing order from left to right.)

C. FAAM-RETRIEVED AND ATLID-MEASURED CLOUD EXTINCTION COEFFICIENT COMPARISON

This section presents aircraft–satellite cloud extinction coefficient comparison plots for measurement runs where dynamic cloud conditions or limited spatial and temporal overlap restrict meaningful quantitative analysis, along with the B895 retrieval plot.

Figure 9 shows the cloud extinction profiles retrieved from flight B895. Retrieval results for this dataset have been cross-validated against previous studies.

Figures 10 and 11 present the cloud extinction profiles retrieved from flights C395 and C397, respectively. These datasets do not contain valid measurements in regions of direct overlap with the satellite observations. Furthermore, rapid cloud evolution during these flights precluded meaningful quantitative comparison with EarthCARE measurements.

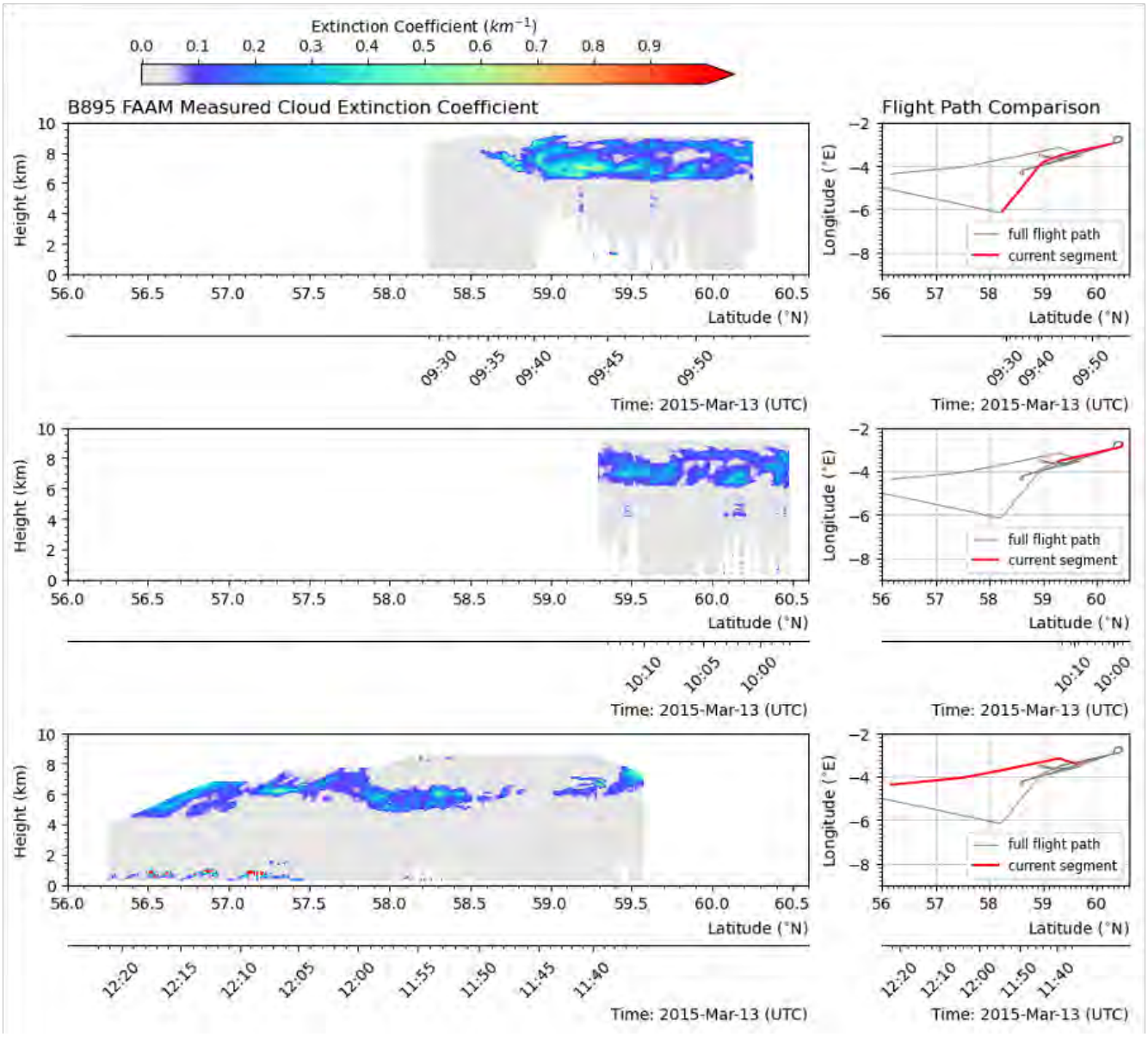

FIG. 9. Cloud extinction profile retrieved from FAAM airborne lidar measurement B895 done on 13th March 2015. The measurement is segmented by the aircraft's heading angle and selected segments are displayed in the figure. The subplots on the left depict the segment's retrieved cloud extinction coefficient profile at latitude and UTC time at which the lidar measurements are taken. The subplots on the right display the flight segment's flight path relative to the aircraft's overall flight path as a reference.

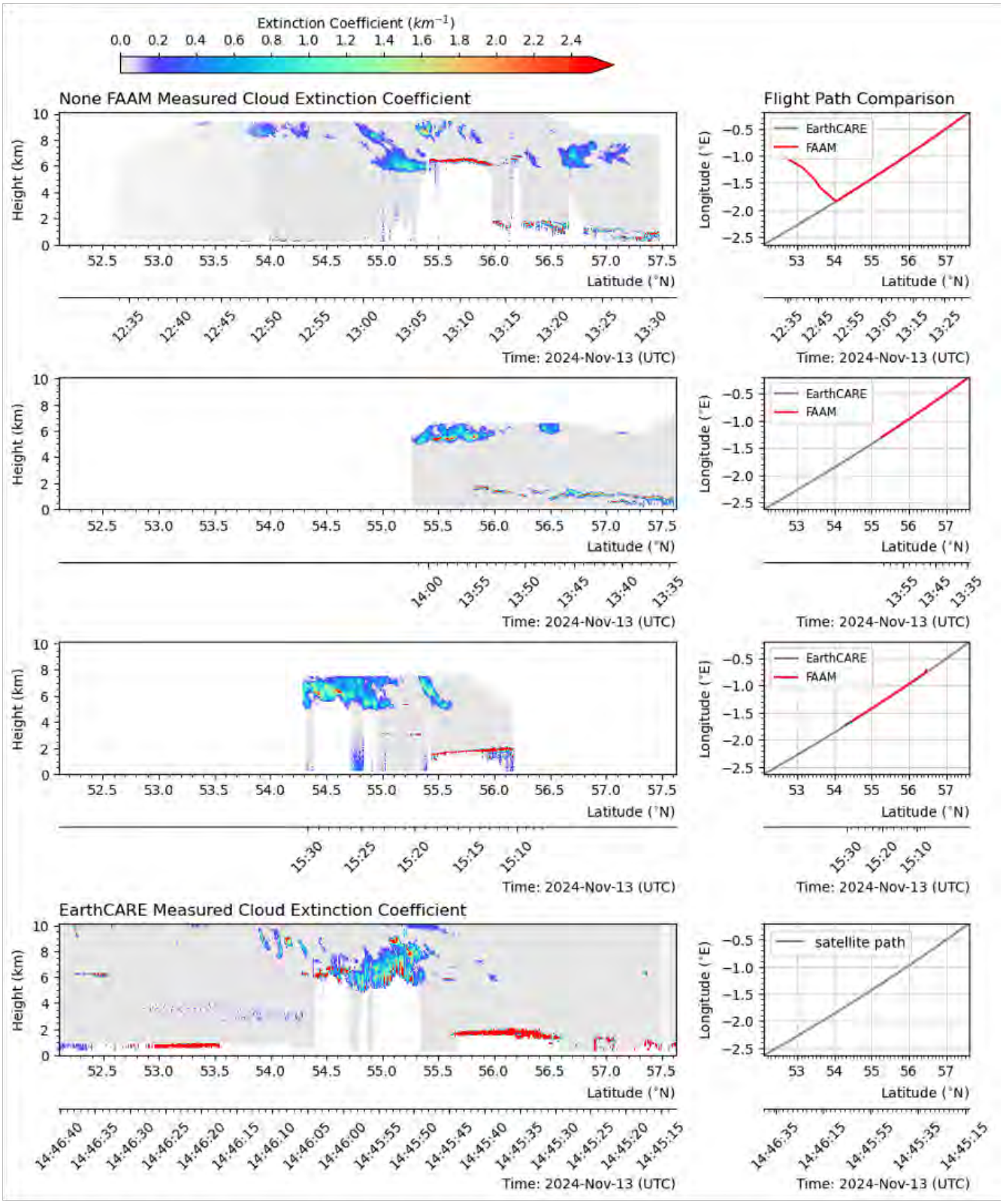

FIG. 10. Comparison of cloud extinction profile retrieved from FAAM airborne lidar measurement C395 and measured by ATLID on EarthCARE on 13th November 2024. Refer to the caption of figure 3 for detailed breakdown of the subplots and features. The C395 measurement does not contain meaningful data in its spatial and temporal overlap region with the satellite.

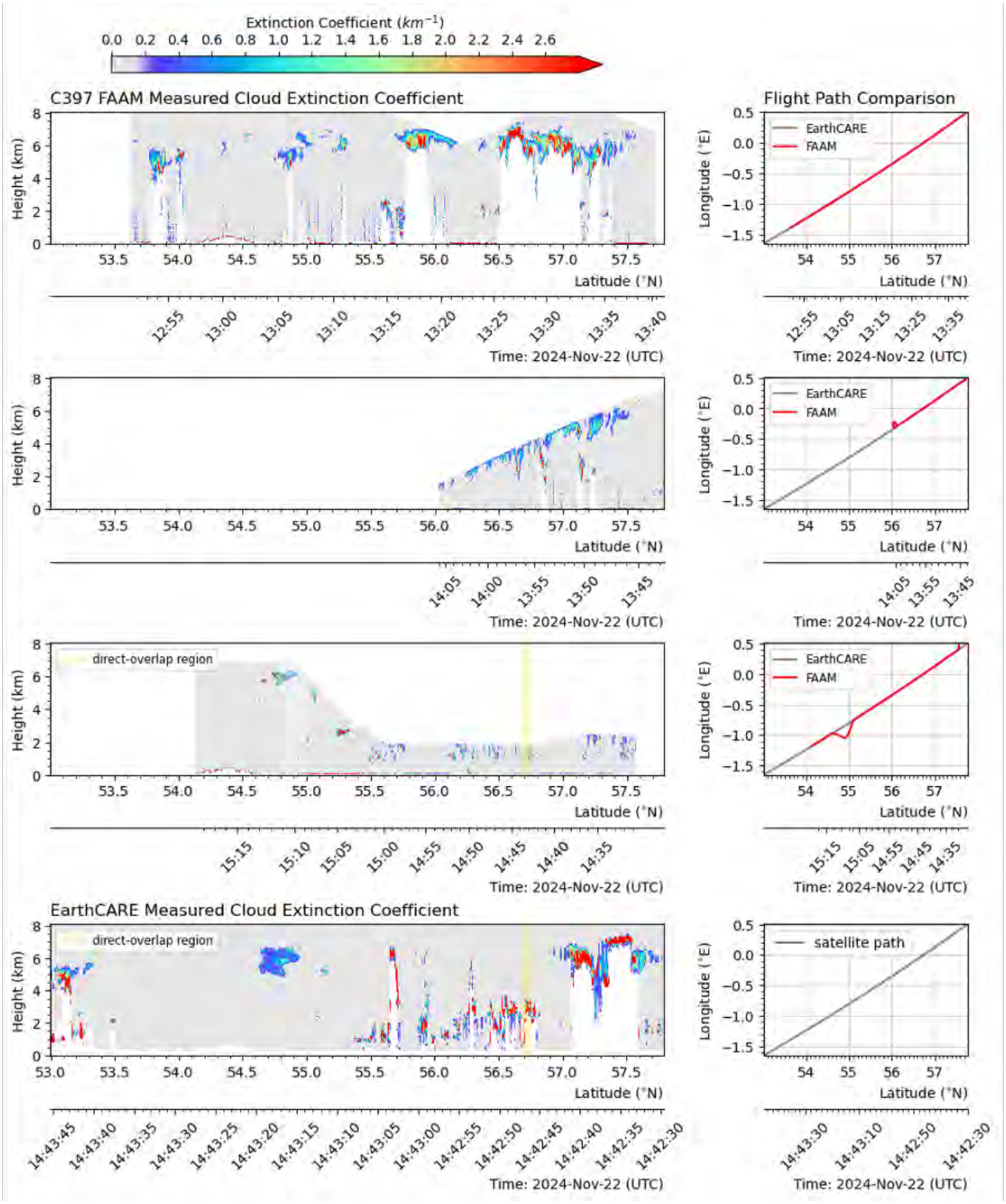

FIG. 11. Comparison of cloud extinction profile retrieved from FAAM airborne lidar measurement C397 and measured by ATLID on EarthCARE on 22nd November 2024. Refer to the caption of figure 3 for detailed breakdown of the subplots and features.