https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

Characterization of CHARM-F Lidar Signals from CoMet Campaigns over Wetlands and Water

Arunima Das[1,2,3], Christoph Kiemle[2,3], Anna Weber[1], Sabrina Zechlau[1,2], Mathieu Quatrevalet[2], and Andreas Fix[2]

[1]Meteorologisches Institut, Ludwig-Maximilians-Universität München, Munich, Germany
[2]Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[3]Technical University of Munich, Department of Aerospace and Geodesy, Munich, Germany

Correspondence: Arunima Das (arunima.das@dlr.de)

Abstract. Methane (CH_4) is the second most important greenhouse gas with an atmospheric lifetime of approximately 10 years. Due to its relatively short lifetime, it offers substantial mitigation potential with reduction in emissions. Lakes and wetlands constitute major natural CH_4 sources, and offshore oil and gas production is a dominant anthropogenic source over open water. However, passive remote sensors have observational gaps over water due to low surface reflectance and challenging sun-glint measurement geometry, and in-situ observations also remain sparse. Active remote sensing using CO_2 and CH_4 Atmospheric Remote Monitoring – Flugzeug (CHARM-F) Integrated Path Differential Absorption (IPDA) lidar provides column-averaged methane mixing ratio (XCH_4) over water bodies and wetlands independent of solar illumination and largely independent of surface reflectance. However, understanding how the lidar signal to noise ratio (SNR) varies with different surface types, and how this variability propagates into retrieval uncertainties is critical to fully exploit the potential of IPDA lidar over water bodies and wetlands.

In this study, we analyze airborne CHARM-F IPDA lidar observations acquired during two Carbon dioxide and Methane (CoMet) field campaigns over heterogeneous land and water surfaces in Europe and North America. This lidar serves as the airborne demonstrator for the future space borne MEthane Remote sensing Lidar missioN (MERLIN) which will deliver global methane emission maps and reduce uncertainties in methane emission estimates. Here, we study the averaging biases of the lidar XCH_4 retrievals and propagate SNR-derived uncertainties to the final XCH_4 estimates building on the work of Tellier et al. (2018). An independent analysis of surface reflectance measurements from the hyperspectral imaging spectrometer of the Munich Aerosol Cloud Scanner (specMACS) against CHARM-F signal strength and SNR provides a direct observational link between surface properties, signal statistics, and retrieval performance. The lidar signals over dry land exhibit Gaussian distributions associated with high surface reflectance, high SNR and low non-correctable residual XCH_4 uncertainties below 1 ppb. In contrast, wetlands and open water surfaces display weaker and broader signal distributions with SNR reduced by a factor of about 2 as compared to dry land. However, the retrieval uncertainty over water bodies remains below 1 ppb when using an appropriate water mask and SNR-weighted averaging of the lidar signals. These findings provide the first observationally-constrained characterization of how surface type is related to lidar SNR, averaging bias, and retrieval uncertainty in airborne IPDA methane measurements. The results highlight the capability of IPDA to robustly measure methane over water, relevant for the MERLIN mission.

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

1 Introduction

Methane (CH_4) is the second most significant anthropogenic greenhouse gas, contributing ∼30% to global warming since pre-industrial times despite its comparatively small share of total carbon emissions (Mwanake et al. (2023), Boy et al. (2022)). Its relatively short atmospheric lifetime (∼10 years) implies a high mitigation potential as compared to CO_2 (Saunois et al. (2025)) yet, substantial uncertainties remain in the global methane budget. In particular, discrepancies between bottom-up inventories and top-down atmospheric inversions indicate incomplete understanding of source strengths and underlying processes (Shindell et al. (2024)).

A large source of uncertainty comes from CH_4 estimates over inland and open water bodies. Inland water and natural wetlands constitute the largest single source of methane, contributing about 30% of global emissions (Saunois et al. (2016)). Their fluxes are highly heterogeneous and controlled by hydrological and bio-geochemical processes, with contributions from diffusion, plant-mediated transport, and ebullition, which vary strongly in space and time (Grinham et al. (2018)). Calculating wetland emissions typically requires simulations from many complex "bottom-up" physics-based models combined with "top-down" atmospheric satellite observations (Li et al. (2026)). These characteristics, combined with limited accessibility but high global fluxes, make wetlands particularly challenging to quantify (McNicol et al. (2023)). Additionally, recent acceleration of methane growth rates has been strongly linked to major wetland regions and climate anomalies, while high-latitude systems add further uncertainty due to permafrost thaw and emerging water (Lin et al. (2024), Kort et al. (2012)). The Arctic Methane and Permafrost Challenge (AMPAC) study highlights how upcoming satellite missions with lidar systems, such as MERLIN, can significantly improve Arctic methane monitoring by reducing uncertainties in permafrost-related greenhouse gas emissions through enhanced large-scale atmospheric observations (Bartsch et al. (2025)). A second key uncertainty comes from offshore oil and gas production, which represents an important anthropogenic source. However, methane emission flux from offshore oil and gas is poorly quantified, primarily due to observational limitations over ocean surfaces with low reflectance (Riddick et al. (2025), Barkley et al. (2023)). In both cases of inland water bodies including wetlands and open ocean surfaces, passive satellite retrievals are strongly constrained by low surface reflectance, cloud cover, and illumination conditions, leading to sparse coverage and potential biases in these regions of high uncertainty over wetlands and water. The reliance of passive satellite retrievals on the sun glint effect over water inherently limits their global coverage primarily to confined regions in the tropics and subtropics (Bréon and Henriot (2006), Lorente et al. (2022), MacLean et al. (2024)). Hence, most offshore methane emission data rely on bottom-up inventories that use generic emission factors, which typically fail to account for "ultra-emitters" - sporadic, large-scale releases caused by equipment failure or maintenance (Fiehn et al. (2025), Riddick et al. (2025)). As a result, both natural aquatic systems and offshore anthropogenic sources remain key gaps in the global methane budget.

Improved remote sensing and satellite observations of atmospheric CH_4 over low-albedo and heterogeneous surfaces are therefore required to reduce uncertainties in methane emissions estimations. Passive remote sensing techniques have greatly improved our knowledge but are also limited by their dependence on solar back scatter, particularly over water bodies and at high latitudes. Active remote sensing using Integrated Path Differential Absorption (IPDA) lidar provides an alternative ap-

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

proach, enabling column measurements of CH_4 largely independent of solar illumination and surface reflectance. Additionally, it also enables measurements in regions with broken clouds and in-between deep tropical convection. This measurement principle forms the basis of the MERLIN mission, developed jointly by Centre National d'Études Spatiales (CNES) and Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), which will deliver global observations of column-averaged methane mixing ratios (XCH_4) (Ehret et al. (2017)).

Achieving the required accuracy with MERLIN measurements over heterogeneous and low-reflectance surfaces depends critically on understanding measurement uncertainties and retrieval biases, particularly those introduced during along-track averaging of the lidar signals. A previous simulation study (Tellier et al. (2018)) has demonstrated the presence of averaging-induced biases arising from the non-linearity of the retrieval and the heterogeneity of the observed scene. However, these effects have not yet been comprehensively evaluated using airborne measurements under realistic observational conditions. Thus, such an assessment is required to support the development and validation of retrieval algorithms for MERLIN (Tellier et al. (2018)).

To this end, airborne campaigns (HALO CoMet campaign (2018), HALO CoMet 2.0 Arctic campaign (2022)) using the CHARM-F IPDA lidar provide a unique testbed for evaluating lidar retrieval performance under realistic conditions. These campaigns include coordinated measurements over diverse surface types. During the CoMet 2.0 Arctic campaign the instrument setup was additionally complemented by the hyperspectral imaging system specMACS (spectrometer of the Munich Aerosol Cloud Scanner) (Ewald et al. (2016)). specMACS delivers high-resolution spectral and spatial information on surface reflectance, enabling a physically consistent characterization of surface properties that directly influence lidar signal strength and retrieval quality. The combined dataset enables an analysis of the relationship between surface properties, lidar signal strength, and retrieval performance across different surface types, including dry land, wetlands, and open water.

In this study, we use observations from the CoMet 1.0 and CoMet 2.0 Arctic campaigns to systematically investigate averaging-induced biases in methane retrievals over heterogeneous surfaces, with a particular focus on water bodies and wetlands. By combining CHARM-F lidar measurements with specMACS surface reflectance and auxiliary water-mask datasets (Pekel et al. (2016), CLMS-Water-Bodies (2022)), we assess how signal strength, surface properties, and averaging strategies interact to produce potential biases. In addition, the relationship between lidar signal strength, signal to noise ratio (SNR), and surface reflectance is investigated to characterize the lidar signals over different surface types.

This work provides the first observationally constrained evaluation of the averaging biases across a wide range of surface conditions and identifies regimes where retrieval performance can be further improved. The significance of this study lies in its direct relevance for the MERLIN mission and future active remote sensing systems. By estimating and improving these averaging biases, dependent on surface heterogeneity and signal-to-noise conditions, we provide guidance for optimal averaging strategies, averaging window lengths, and uncertainty characterization. By addressing key limitations in current methane observations over low-reflectance surfaces, this work helps to reduce uncertainties in methane emission estimates from both natural and anthropogenic sources.

The paper is organized as follows. Section 2 gives an overview of the CHARM-F lidar and the specMACS hyperspectral imager. Section 3 describes the averaging schemes and bias calculation procedures from Tellier et al. (2018) along with an

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

uncertainty propagation formulation, and an overview of the two CoMet campaigns is given in Section 4. Finally, in Section 5, the CHARM-F signals are characterized and compared against specMACS surface reflectances, and the averaging biases over land, water, and wetlands are quantified. The paper wraps up with limitations in the current work and implications for the MERLIN mission.

2 Instruments Overview

2.1 IPDA Measurement Principle

The CHARM-F Integrated Path Differential Absorption (IPDA) lidar is in preparation for the validation of MERLIN providing precise measurement of the column-averaged dry-air mixing ratio of atmospheric methane (XCH_4) and is capable to measure also atmospheric carbon dioxide (XCO_2) in the same manner (Amediek et al. (2017)). CHARM-F is an IPDA-lidar-based remote sensing method designed to use hard target reflections from the Earth's surface in the near infrared (NIR) region to measure both greenhouse gases with its typical laser footprint of $10\,\mathrm{m}$ at $10\,\mathrm{km}$ flight altitude. It provides a direct measurement of the atmospheric path using the runtime measurement and is weakly influenced by aerosol and thin clouds in comparison to passive remote sensors.

The two lidar systems for CH_4 and CO_2 use pulsed direct detection at $1645\,\mathrm{nm}$ for CH_4 and $1572\,\mathrm{nm}$ for CO_2, backscattered from a "hard target" namely ground, water surfaces, or the top of a cloud layer. CHARM-F is based on two narrow-band optical parametric oscillators pumped by Q-switched Nd:YAG lasers in a master-oscillator power-amplifier configuration. A key technical requirement is the highly precise stabilization of the laser wavelengths, which is achieved using a multi pass gas cell filled with both CO_2 and CH_4 as an absolute wavelength reference (Fix et al. (2011)).

The IPDA technique relies on the differential absorption of laser light by the target gas over a well-defined atmospheric column. The lidar emits pulse pairs per trace gas at slightly different wavelengths named online and offline (Ehret et al. (2008)). The online wavelength is positioned in the vicinity of a CO_2 or CH_4 absorption line in the NIR spectral region. For CH_4, the online pulse at $1645.552\,\mathrm{nm}$ experiences significant molecular absorption during its propagation through the atmosphere. Another measurement with negligible absorption by the target gas is denoted offline ($1645.846\,\mathrm{nm}$) and serves as reference wavelength located far enough from the online wavelength. At the same time, the on- and offline wavelengths are spectrally close enough to consider the atmospheric and surface properties to be identical with the exception of the greenhouse gas absorption. Such an online-offline pulse pair measurement is referred to as a "single-shot measurement". CHARM-F transmits these online and offline pulses as a double-pulse pair with a short temporal separation of $500\,\mu\mathrm{s}$ corresponding to $0.1\,\mathrm{m}$ on ground for HALO research aircraft (Schmidt et al. (2018)) typically flying at $200\,\mathrm{ms}^{-1}$ and $0.05\,\mathrm{m}$ on ground for ATR-42 aircraft (SAFIRE ATR-42 aircraft(2025)) typically flying at $100\,\mathrm{ms}^{-1}$. CHARM-F has a larger telescope ($200\,\mathrm{mm}$ diameter) with a PIN photo diode as the detector and a smaller telescope ($60\,\mathrm{mm}$ diameter) with an avalanche photo diode (APD), which allows for comparison of the characteristics of these two detector types under real operation conditions. While CHARM-F measures both greenhouse gases, CH_4 and CO_2, our work here focuses on the CH_4 measurements.

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

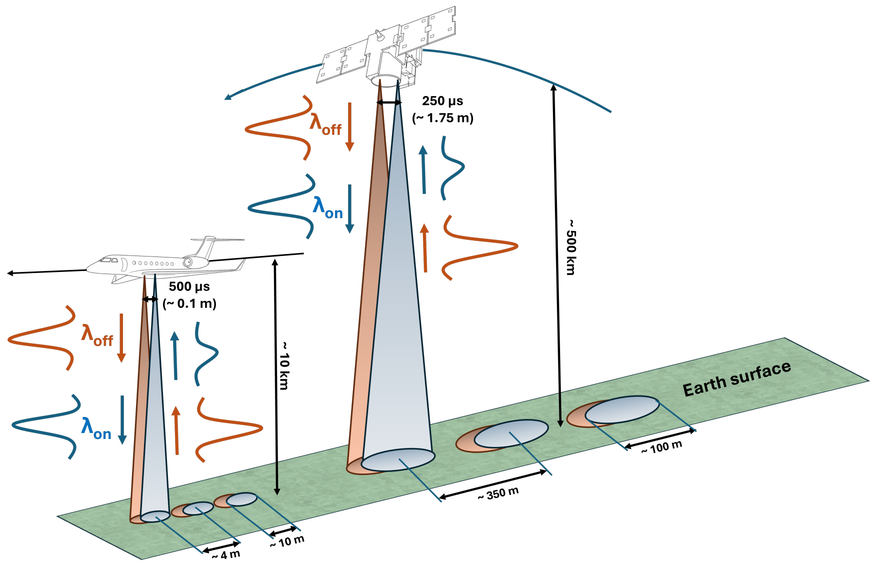


Figure 1. Schematic showing the measurement principle and geometry for the lidar measurements from an aircraft on the left and satellite platform on the right. The alternate online and offline pulses are indicated in blue and red colours respectively. The Earth surface which serves as the hard target for the return pulses is shown in green texture. The different parameters for the measurements including the altitude, footprint sizes, pulse separation times are also mentioned. Note that for visualization purpose, the distance between the on- and offline footprints is exaggerated as compared to the footprint diameter.

Figure 1 shows the measurement principle and the geometry for the lidar from an aircraft at an approximate altitude of 10 km and for a low Earth orbit satellite at an altitude of 500 km. Table 1 provides the system specifications and operational flight parameters for the HALO (airborne) and MERLIN (space borne) lidar instruments (MERLIN Satellite Mission(2025)). The metrics highlight the scaling differences in footprint size, pulse timing, and receiver apertures required when transitioning from an aircraft platform to a low Earth orbit satellite.

The integrated column concentration is derived using a differential absorption of the online and offline lidar echoes along with the determined range between the lidar and back scattering target. This gives the Differential Absorption Optical Depth (DAOD) based on Eq. 1.

$$DAOD = \frac{1}{2} \cdot \ln\left(\frac{S_{off}^{target} / E_{off}^{ref}}{S_{on}^{target} / E_{on}^{ref}}\right) \quad (1)$$

S_{on}^{target} and S_{off}^{target} are the energy scattered back from the target for online and offline laser pulses, respectively. For each laser pulse, a corresponding energy reference measurement E^{ref} is performed inside the lidar instrument with the help of a beam splitter that deflects a small portion of its energy onto a detector for each individual laser pulse (Fix et al. (2018)). In this way, the laser pulse energies E_{on} and E_{off} are determined separately as a reference within the lidar system.

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

Table 1. Difference between the Measurement Geometry and Parameters for Lidar from the HALO Aircraft and MERLIN Satellite

Parameter	**HALO**	**MERLIN**
Altitude	10 km	500 km
Footprint Diameter on ground	10 m	100 m
Online-Offline Pulse Separation (Corresponding distance on ground)	500 μs (0.1 m)	250 μs (1.75 m)
Pulse Pair repetition rate (Corresponding distance on ground)	50 Hz (4 m)	20 Hz (350 m)
Receiving Telescope size	200 mm (PIN)	69 cm
	60 mm (APD)	
Detector Type	PIN and APD	APD
Speed	200 ms^{-1}	7000 ms^{-1}

XCH_4 is calculated by weighting the DAOD with the CHARM-F weighting function (WF) as per Eq. 2 (Ehret et al. (2008), Krautwurst et al. (2025)). A description on the derivation of the WF can be found in the work of Ehret et al. (2008).

$$X_{CH_4} = \frac{DAOD}{\int_{p_{aircraft}}^{p_{target}} WF(p,T) \cdot dp} \tag{2}$$

The integral of the weighting function describes the atmospheric column mass and spectroscopic sensitivity, and is determined using external data, such as, molecular spectroscopic databases namely, High Resolution Transmission (HITRAN, Rothman et al. (1998)) and atmospheric parameters (pressure, temperature, water vapor) from Numerical Weather Prediction (NWP) models. The target pressure (p_{target}) is determined for each single-shot using the high-precision ranging capability of the lidar (Amediek et al. (2017)).

The single-shot version of Eq. 2, giving $X_{CH_4,i}$, is denoted by Eq. 3 which is calculated in the same way with δ_i representing the single shot DAOD and IWF_i representing the Integrated Weighting Function for each shot.

$$X_{CH_4,i} = \frac{\delta_i}{IWF_i} \tag{3}$$

The hard target reflections give a higher SNR as compared to atmospheric returns, but the surface reflectance and the ground elevation are of particular importance. Both surface reflectance and ground elevation parameters have a significant impact on the DAOD retrieval as they affect the back scattered signal. This in turn affects the mixing ratio retrieval XCH_4, as Eq. 2 indicates, and can vary widely. Also, since the return signals are weak, it is necessary to accumulate several single shot measurements along the flight track in order to achieve the required measurement precision. This forms the basis of the averaging of measurements along the flight track in addition to the integration in the vertical column.

2.2 Hyperspectral and polarized imaging system specMACS

specMACS is a hyperspectral and polarization-resolving imaging system, developed by Ludwig-Maximilians-Universität Munich (Ewald et al., 2016; Weber et al., 2024). During one of the flight campaigns described here, CoMet 2.0 Arctic, the

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

instrument was operated in a downward-looking mode on board the German research aircraft HALO. specMACS comprises two hyperspectral cameras for the measurement of solar radiation in the visible and near-infrared (VNIR) and the shortwave-infrared (SWIR) spectral ranges, covering a total wavelength range of 400 to 2500 nm with a nominal spectral resolution of 3 nm and 10 nm for the VNIR and SWIR, respectively. The hyperspectral cameras of the specMACS instrument are push broom line imagers and operate by projecting a slit onto a diffraction grating which provides spectral separation of the incoming light to the image sensors. They have a field of view of 32.7 ° and 35.5 ° in across-track direction, respectively, and a spatial resolution on the order of 10 m for objects at a distance of about 10 km and a typical data acquisition rate of 30 Hz. The overall radiometric uncertainty of the VNIR and SWIR cameras is around 5 % and 10 %, respectively. Besides the spectrometers, specMACS includes two 2D RGB polarization-resolving cameras, which are, however, not used in this work.

The processing of the specMACS radiance requires three major corrections namely the radiometric, atmospheric, and geometric corrections to convert raw digital numbers into at-sensor radiance values, to remove the effect of the atmosphere and convert at-sensor radiance to at-surface reflectance, and finally to compensate image distortions caused by the aircraft movements, orthogonalize data, and bring the data into a defined coordinate system. The radiometric calibration and the georeferencing of the data were performed as described in Ewald et al. (2016) and Weber et al. (2024). For the computation of the surface elevation of every measured pixel for the atmospheric correction, the Copernicus GLO-30 Digital Elevation Model (Copernicus Data Space Ecosystem, 2021) was applied. The atmospheric correction itself was then performed based on the Atmospheric Correction Software for Airborne and Satellite Imaging (ATCOR-4, Richter, 1996; Richter et al., 2006; Richter and Schläpfer, 2016). The final atmospherically corrected data product represents the spectral hemispherical directional reflectance factor, for simplicity called surface reflectance in the following. It provides information about the reflecting properties of the surface based on a physical quantity and is applied for a comparison with the CHARM-F offline back scatter signal. The basic viewing geometries of specMACS and CHARM-F are slightly different and displayed in Fig. 2. The passive remote sensing instrument specMACS measures incoming sunlight, which is reflected into the direction of the sensor, whereas CHARM-F is an active remote sensing instrument emitting and receiving reflected radiation in the nadir and zenith directions.

3 Averaging Schemes and Bias Characterization

To balance the trade-offs between random detector noise and non-linear systematic biases, the literature (Tellier et al. (2018)) mentions three different averaging schemes for IPDA measurements, namely Averaging of Signals (AVS) to maximize SNR in weak signal regimes, Averaging of Optical Depths (AVD) to maintain linearity with atmospheric column mass, and Averaging of Mixing Ratios (AVX) for direct geophysical application. Tellier et al. (2018) considered AVS the most robust approach for minimizing the total systematic error which constrains the accuracy of the retrieved dry-air methane mixing ratio (XCH_4) in simulated flight conditions during horizontal averaging of the lidar data. The present study uses measured data to corroborate this result. The methodology for the comparison of the averaging approaches along with bias calculation follows the one as developed by Tellier et al. (2018). The focus of this study is on the characterization of the lidar signals over different surface types with an assessment of the non-correctable uncertainty on the XCH_4 estimate but without a full assessment of detector

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

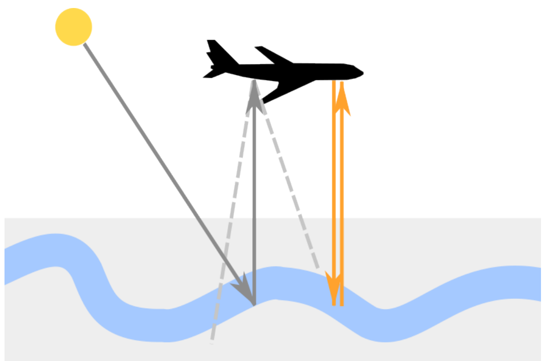


Figure 2. Basic viewing geometry of the specMACS instrument (gray lines) in comparison to CHARM-F (orange lines). The gray dashed lines indicate the field of view of the spectrometers of specMACS in across-track direction and the yellow circle represents the sun. The gray and blue shaded areas indicate land and water surfaces.

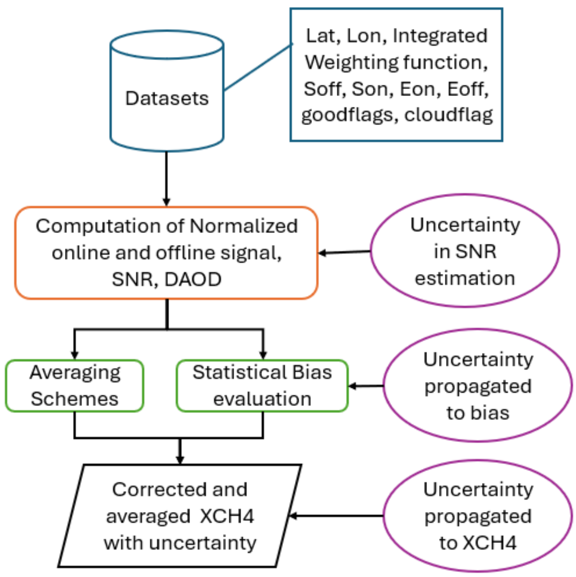


Figure 3. Workflow for applying the Averaging schemes on the CHARM-F measurements along with the Bias Calculation and Uncertainty Propagation on the final XCH_4 estimate. It shows the overall steps from the inclusion of the datasets to the SNR calculations and the averaged XCH_4 estimation with the propagated uncertainty.

nonlinearities and/or saturation effects which is out of the scope of this paper. Additionally, in this work we use airborne lidar measurements obtained during two flight campaigns instead of a simulated measurement setup. This helps us to critically analyse the bias and uncertainty propagation over heterogeneous surface returns using airborne observational data. The overall workflow for this analysis is shown in Figure 3.

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

We apply the averaging schemes methodology for a defined window of 151 shots (corresponding to 3 s) of CHARM-F measurements over dry land, wetland and open water surfaces. In equations 4, 5, and 6 under the following subsections, the notation $\langle X \rangle$ denotes the averaging of the quantity X across the chosen averaging window. Each approach represents a different
200 order of non-linear operations from online and offline signals to XCH_4 column concentration, and therefore, gives different averaging biases. However, over dark surfaces like water having low surface reflectance, the noise may cause the online or offline signal to drop below zero. Since the logarithm of a negative number is undefined, these shots must be discarded and this, therefore, creates a systematic bias.

3.1 AVX - Averaging of Mixing Ratios

205 AVX computes XCH_4 independently for each online–offline shot pair and then averages these retrieved values to obtain a mean value over the desired averaging window as shown in Equation 4. The averaging is IWF-weighted (denoted by w[IWF]), meaning shots with larger IWF values contribute more to the final mean. Because XCH_4 is a nonlinear function of the measured signals, each single-shot retrieval is biased by statistical noise. However, averaging does not cancel these biases, but accumulates biases from individual shots instead. This effect becomes especially severe over low-reflectance surfaces such as
210 water, where the SNR is low and noise-induced fluctuations are present in the online and offline signals.

$$\overline{X_{CH_4}}^{AVX} = \left\langle \frac{\delta}{IWF} \right\rangle w[IWF] \tag{4}$$

3.2 AVD – Averaging of DAOD retrievals

The second method, AVD, seeks to mitigate some of these issues by computing the DAOD for each shot pair, averaging the DAODs, and then converting the averaged DAOD to XCH_4 as shown in Equation 5. This averages the extensive properties of
215 the measurements separately and reduces the degree of nonlinearity compared to AVX. Nevertheless, the logarithmic transformation used to compute the DAOD from noisy signals as shown in Equation 1 still introduces an averaging bias because the logarithmic transformation is itself nonlinear. Therefore, even if the online and offline signals individually follow symmetric Gaussian distributions, the computed DAOD will follow a skewed distribution due to the logarithm. Additionally, the mean of this skewed distribution is not equal to the true DAOD. Moreover, if the underlying scene is heterogeneous, with significant
220 variation in surface reflectance across the averaging window, AVD can still be biased because all DAODs are weighted equally regardless of their signal strength.

$$\overline{X_{CH_4}}^{AVD} = \frac{\langle \delta \rangle}{\langle IWF \rangle} \tag{5}$$

3.3 AVS – Averaging of signals

The AVS scheme performs the averaging at the raw signal level before the non-linear logarithmic transformation is applied.
225 This method is particularly effective for minimizing averaging bias in low reflectance and offline signal strength conditions. Specifically, the mean online and offline signals across all shots within the averaging window, here denoted as Q, are calculated before the DAOD is derived from their ratio (see Equation 1). Because the averaging occurs prior to the nonlinear operations,

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

the averaging bias is greatly reduced. Furthermore, AVS implements signal-based weighting, i.e. brighter surfaces contribute more strongly to the averaged signal, while dark surfaces contribute less. This makes the AVS scheme more robust to albedo variability across the averaging window. For these reasons, AVS emerged in the MERLIN study as the preferred scheme (Tellier et al. (2018)).

$$\overline{X_{CH_4}}^{AVS} = \frac{\frac{1}{2} ln\left(\frac{Q^{off}}{Q^{on}}\right)}{\langle IWF\rangle_{w[Q^{off}]}} \tag{6}$$

3.4 Bias Estimation

When averaging data over a window, the bias arises from the interaction between the instrument's detector noise and the non-linear logarithm function used to calculate the DAOD. Both AVX and AVD schemes calculate the DAOD for every single shot pair before averaging. Therefore, even if the online and offline signals are symmetrical Gaussian, the calculated DAOD turns out to be a skewed distribution due to the non-linear logarithmic relation (Tellier et al. (2018)).

Tellier et al. (2018) derived the equations to correct the averaging bias by integrating a "truncated normal distribution" as shown in Eq. 7 and 8 which can also be referred to as the integral bias.

$$Bias_{stat}(\hat{\delta}) = \mathrm{b} = \frac{1}{2} \cdot E\left[\ln\left(1 + \frac{X^{off}}{SNR^{off}}\right)\right] - \frac{1}{2} \cdot E\left[\ln\left(1 + \frac{X^{on}}{SNR^{on}}\right)\right] \tag{7}$$

where,

$$E\left[\ln\left(1 + \frac{X^{on,off}}{SNR^{on,off}}\right)\right] = \frac{1}{\sqrt{2\pi}\left[1 - \Phi\left(-SNR^{on,off}\right)\right]} \int\limits_{-SNR^{on,off}}^{+\infty} \ln\left(1 + \frac{x}{SNR^{on,off}}\right) \cdot e^{-\frac{x^2}{2}}\, dx \tag{8}$$

When averaging the raw online and offline signals across a window of N shots with AVS, the weighting of the retrieval is directly proportional to the signal strength. The reason for this is rooted in the physics of how the lidar signal is formed. The back scattered signal received by the CHARM-F detector from each laser shot is directly proportional to the surface reflectance at that moment. When the signals are averaged, the bright surface shots contribute more to the average value than the dark surface shots and, therefore, averaging raw signals inherently gives more weight to laser shots reflected from brighter surfaces with higher reflectance. To account for this, the IWF used to derive CH_4 is not averaged uniformly but is weighted by the offline signal strength (Q^{off}) of each shot. The selection of an averaging scheme, therefore, involves a trade-off between sensitivity to measurement noise and sensitivity to atmospheric variability. Both the AVX and AVD schemes preserve the shot-by-shot geophysical information but are highly susceptible to the bias caused by the non-linearity of the retrieval equation at low SNR. Therefore, both methods require complex bias corrections. In contrast, it is seen that the AVS scheme effectively eliminates averaging bias for simulated noisy measurements by boosting the effective SNR through signal summation prior to the logarithmic processing. This makes the AVS method robust even over dark targets (Tellier et al. (2018)).

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

3.5 Uncertainty Propagation

Tellier et al. (2018) correct methane column retrievals for biases that arise from nonlinear transformations of noisy online and offline signals. These corrections depend explicitly on the SNR estimation of the online and offline channels, as we can see from Equations 7 and 8. Consequently, uncertainties in the estimated SNR propagate directly into the corrected DAOD and ultimately into XCH_4. Therefore, to quantify the robustness of the bias correction, we propagate SNR uncertainties into an uncertainty on the corrected XCH_4 using first-order error propagation.

3.5.1 Uncertainty Estimation on SNR

The signal to noise ratio for online and offline signals is defined as in Equation 9.

$$\mathrm{SNR} = \frac{\mu}{\sigma} \tag{9}$$

where μ is the sample mean and σ is sample standard deviation. The SNR estimation has an uncertainty which can be written assuming a Gaussian noise distribution. For N independent samples with standard deviation σ, the uncertainty of sample mean can be written as in Equation 10. The assumption of independence requires that successive shots are uncorrelated, which is reasonable for the 50 Hz CHARM-F repetition frequency over slowly varying surfaces.

$$\sigma_\mu = \frac{\sigma}{\sqrt{N}} \tag{10}$$

The uncertainty of the sample standard deviation is given based on the χ^2-distribution, which to first order approximation can be written as Equation 11.

$$\sigma_\sigma \approx \frac{\sigma}{\sqrt{2(N-1)}} \tag{11}$$

Now, the first order error propagation is used to find the uncertainty on the SNR estimation. This assumes the uncertainties are small and uncorrelated and higher order terms are negligible.

The partial derivatives can be written as in Equations 12 and 13.

$$\frac{\delta SNR}{\delta \mu} = \frac{1}{\sigma} \tag{12}$$

$$\frac{\delta SNR}{\delta \sigma} = \frac{-\mu}{\sigma^2} = \frac{-SNR}{\sigma} \tag{13}$$

These uncertainty values are entered in the first order error propagation formula which then simplifies to Equation 14.

$$\sigma_{SNR}^2 = \left(\frac{1}{\sigma}\right)^2 \sigma_\mu^2 + \left(\frac{SNR}{\sigma}\right)^2 \sigma_\sigma^2 = \frac{1}{N} + \frac{SNR^2}{2(N-1)} \tag{14}$$

The relative uncertainty on the SNR is calculated based on Equation 15.

$$\frac{\sigma_{SNR}}{SNR} = \sqrt{\frac{1}{N} + \frac{1}{2(N-1)}} \tag{15}$$

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

In summary, the SNR uncertainty is derived from the sampling statistics of the averaged signal using first-order error propagation, accounting for uncertainties in both the sample mean and standard deviation of the signal. For the assumed Gaussian distribution, the resulting relative SNR uncertainty depends only on the number of samples.

3.5.2 Uncertainty Propagation on DAOD and XCH4

This SNR uncertainty is subsequently propagated into the statistical bias correction using first-order Taylor expansion of the bias expression with respect to SNR coming directly from Equation 7. Finally, the bias uncertainty is propagated linearly to XCH_4 through division by the integrated weighting function. The partial derivatives of the averaging bias with respect to the online and offline SNR can be written in discretized form as in Equations 16 and 17, respectively, using a centered finite-difference approximation. b is the statistical bias correction value in DAOD units, evaluated at the estimated SNR values for online and offline channels. SNR_{on} and SNR_{off} are the estimated signal-to-noise ratios for the online and offline channels, respectively, computed from the averaged signal statistics within the window using Equation 9. ϵ is a small numerical perturbation chosen to be large enough to produce a numerically stable finite difference but small enough that the linear approximation of the Taylor expansion remains valid. In practice ϵ is chosen as the uncertainty of the SNR estimation using Equation 15.

$$\frac{\delta b}{\delta SNR_{on}} \approx \frac{b(SNR_{on}+\epsilon, SNR_{off}) - b(SNR_{on}-\epsilon, SNR_{off})}{2\epsilon} \tag{16}$$

$$\frac{\delta b}{\delta SNR_{off}} \approx \frac{b(SNR_{on}, SNR_{off}+\epsilon) - b(SNR_{on}, SNR_{off}-\epsilon)}{2\epsilon} \tag{17}$$

The physical meaning of these derivatives is the sensitivity of the bias correction to errors in the SNR estimate. A large $\delta b/\delta SNR_{on}$ means that a small error in the estimated online SNR produces a large change in the computed bias correction and therefore a comparatively large uncertainty in the corrected XCH_4. This sensitivity is highest at low SNR values, where the bias correction itself is changing rapidly with SNR, and lowest at high SNR, where the bias correction is nearly flat and insensitive to small SNR estimation errors. Finally, the uncertainty of the integral bias is given by Equation 18.

$$\sigma_b^2 = \left(\frac{\delta b}{\delta SNR_{on}}\right)^2 \sigma^2_{SNR,on} + \left(\frac{\delta b}{\delta SNR_{off}}\right)^2 \sigma^2_{SNR,off} \tag{18}$$

Since the integrated weighting function does not have an uncertainty due to averaging, for the final uncertainty of XCH_4, we divide the integral bias uncertainty by the IWF as in Equation 19. This is valid because the IWF is treated as exact, being determined from NWP model atmospheric profiles and spectroscopic databases (such as HITRAN), and its uncertainty is not considered within the scope of this study. Only the uncertainty arising from imperfect SNR estimation is propagated here. The division by IWF propagates the bias uncertainty from DAOD space into XCH_4 space. Physically, a larger IWF means that the same DAOD bias uncertainty translates into a smaller XCH_4 uncertainty, because each unit of DAOD corresponds to a higher concentration of methane. Conversely, at high altitude or under weak absorption conditions, the IWF is smaller and the same DAOD uncertainty maps to a larger XCH_4 uncertainty.

$$\sigma_{XCH_4} = \frac{\sigma_b}{\mathrm{IWF}} \tag{19}$$

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

4 Airborne Methane Lidar Measurements - CoMet Flight Campaigns

To prepare and validate the MERLIN mission, methane column measurements were undertaken by CHARM-F onboard the German HALO research aircraft measuring CO_2 and CH_4 during various measurement campaigns (HALO CoMet campaign (2018), HALO CoMet 2.0 Arctic campaign (2022)). These airborne campaigns help to improve our current understanding of the distribution of CO_2 and CH_4. In this context the following field campaigns were executed over dedicated sources in Europe and Canada with the CHARM-F instrument onboard the German High Altitude and Long Range (HALO) research aircraft. This section provides an overview of the CoMet 1.0 and CoMet 2.0 Arctic measurement campaigns whose data we utilize to evaluate the biases and SNR over land and water.

CoMet 1.0 took place during May and June 2018, with a focus on anthropogenic greenhouse gas emissions over Europe. It aimed at characterizing the distribution of CH_4 and CO_2 over significant regional sources including coal mines and coal-fired power plants, volcanoes, as well as long distance gradients over water surfaces. During the mission, HALO performed nine research flights, with more than 63 hours of observations, with the base of operations located in Oberpfaffenhofen (near Munich, Germany). Each flight aimed to reach several scientific goals based on the meteorological conditions over selected target areas and different flight patterns were executed in order to obtain the most valuable data. Four out of nine flights covered open water surfaces flying over the Baltic, Mediterranean, and North Sea. These flights are particularly relevant for comparing the SNR and biases over land against water surfaces. However, in the CoMet 1.0 campaign, no specMACS or similar instrument was deployed to complement the IPDA lidar signal measurements. Hence, comparison of the signal strength against the specMACS measured surface reflectance can not be carried out.

The second campaign, CoMet 2.0 Arctic, took place over Canada in August and September 2022. The main objective of this campaign was to improve the understanding of the spatiotemporal distribution of methane and carbon dioxide in high Northern latitudes and to compare with existing emission inventories. From the campaign base in Edmonton, Canada, a total of 18 flights were completed in mainly cloud-free conditions, over boreal wetlands, landfills, coal mines, oil and gas mining, permafrost, and forest fires. Several flights targeted the large wetland regions over the Hudson Bay Lowlands, Lake Winnipeg and Mackenzie River Delta. During this campaign, the specMACS instrument was mainly used for the characterization of land surface properties.

Figures 4 and 5 give an overview of all the flights conducted during the CoMet 1.0 and CoMet 2.0 campaigns. For this study, we separate the flights over pure land surfaces, ocean and wetland areas to study and compare the SNR and bias variations over different surfaces. However, the effect of different land surface types (e.g. forest, desert, vegetation covers) are not assessed in this study. In order to mask regions with water we apply an external water-mask dataset, having a 100 m resolution from Copernicus (CLMS-Water-Bodies (2022)) to mask the inland water bodies including lakes, rivers, etc. during the CoMet 2.0 campaign. For measurements from the CoMet 1.0 campaign, we apply a second global surface water-mask from Copernicus, having a 300 m resolution (Pekel et al. (2016)) to mask the ocean and sea surfaces. We consider only ground target returns for this study by using a lidar-derived cloud-mask to discard the signals obtained from cloud return. Therefore, taken together we

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

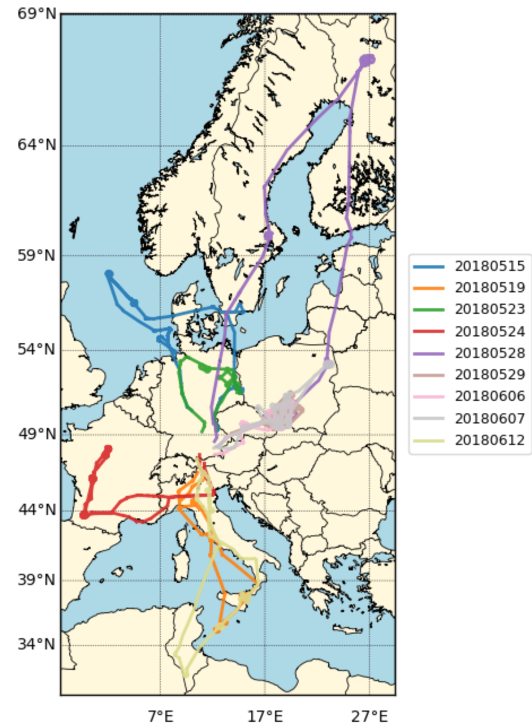


Figure 4. Flight track overview of the CoMet 1.0 campaign, showing all nine scientific research flights across Europe and the Mediterranean. Individual flights are color-coded with the flight date as identifier. The campaign was based in Germany and covered a wide geographic domain, with flight tracks targeting methane source regions including the Upper Silesian Coal Basin (Poland), the North Sea, and the Mediterranean.

have a cloud mask and water mask to process the flights over wetlands and water surfaces and only the cloud mask to process the flights over land from both the campaigns.

Measurements from specMACS as shown in Figure 6 complement the lidar measurements. It gives and example time series of surface reflectance derived from measurements of the SWIR camera of specMACS for different scenes during the three flights over the Mackenzie River Delta, Lake Winnipeg, and Hudson Bay Lowlands, respectively. Water surfaces such as lakes or rivers appear as dark areas with a low surface reflectance due to their low albedo at the considered wavelength in the shortwave infrared wavelength range and the viewing geometry of the instrument.

5 Results

This section shows the major results of the work starting with the CHARM-F offline signal strength characterization over different surface types across the two CoMet flight campaigns. It then gives the outcome of the independent comparison of the specMACS surface reflectance against CHARM-F offline signal strength and the XCH_4 SNR. Finally, this section shows the

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

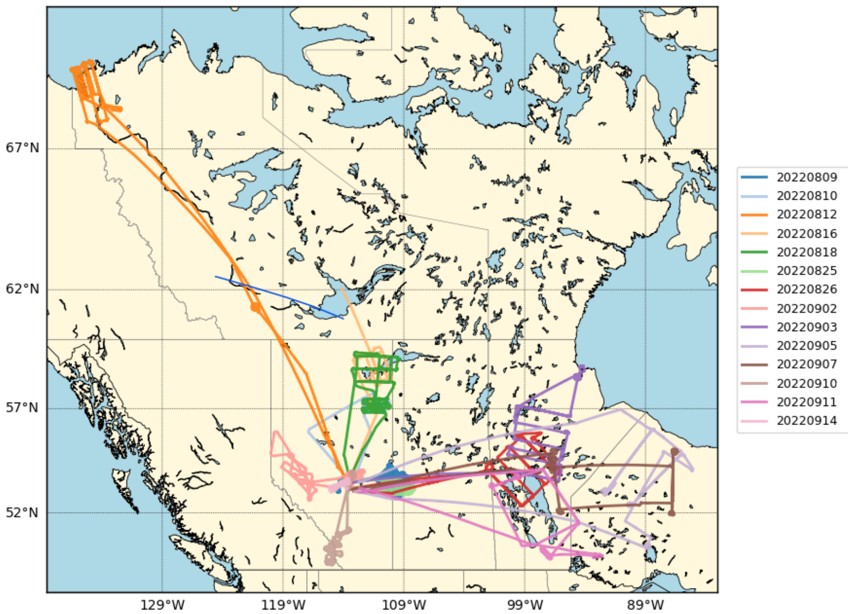


Figure 5. Flight track overview of the CoMet 2.0 Arctic campaign similar to figure 4, showing all scientific research flights over western and central Canada (excluding the transfer flights). The campaign was based in Edmonton, Alberta, and targeted a diverse range of methane source types across the boreal and sub-Arctic regions of Canada including both natural (wetlands, permafrost, wildfires) and anthropogenic (oil sands, oil and gas, coal mines) sources.

statistics of the averaging bias and the non-correctable residual uncertainty on the XCH_4 estimate and how the XCH_4 SNR statistics influences the uncertainty values across different surface types from the two campaigns.

5.1 CHARM-F Offline Signal Characterization

In this section, we evaluate how the CHARM-F offline signal strength varies with changing scenes and surface conditions of different flights from the two campaigns shown as a compilation in Figure 7. The fundamental quantities retrieved for each individual laser shot i are the calibrated online and offline signals, denoted as S_i^{on} and S_i^{off}, respectively. The line-of-sight distance from the lidar sensor to the back scattering target, calculated from the round trip time of the laser pulse is denoted as the range r. The range is critical to isolate the ground return from the atmospheric back scatter or cloud targets and to define the exact column of air being measured for XCH_4. The offline signals are normalized by the square of the range r. The single-shot DAOD, δ_i, is derived from the differential transmission based on Eq. 1. The quantity $S_{off} \cdot r^2$ (range squared normalized offline signal strength), calculated from CHARM-F Level 1 data, is taken as a proxy for the albedo for the surfaces that produced the lidar echo while atmospheric scattering is neglected. It is to be noted that S_{off} is the pure lidar surface return signal without solar background radiation, which also reaches the detector. For all the comparisons and analysis below, the S_{off} has been normalized with the squared range value ($S_{off} \cdot r^2$) such that there is no influence of the range on the compared offline signal

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

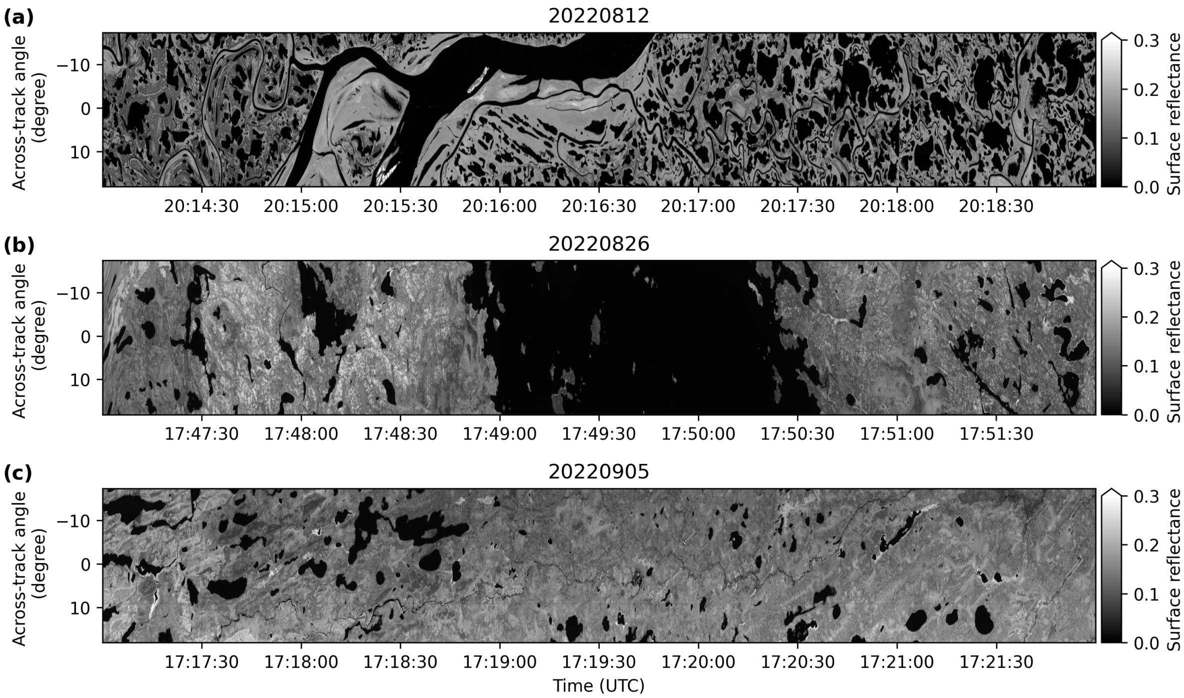


Figure 6. Surface reflectance at 1645 nm wavelength derived from specMACS observations for three example scenes over the Mackenzie River Delta (a), Lake Winnipeg (b), and the Hudson Bay Lowlands (c).

strength across all flights. Some flights from both campaigns focused on emissions over land sources only, while others include flight legs over wetlands, lakes, river deltas and oceans. The variation in the probability density function of the range squared corrected offline signal strength ($S_{off} \cdot r^2$) shows a bimodal distribution with small peaks in the lower reflectance regions for flight segments over wetlands and inland waters (dashed lines in Figure 7). For flights over land sources only, we see a uni-modal close to Gaussian distribution of the signal strength with a pronounced peak at higher reflectance (solid lines in Figure 7). On the other hand, the flights with long remote sensing legs over open oceans have a rather flat and broad distribution of offline signals (dotted lines in Figure 7). The PDFs of the offline signal strengths represent unique signatures for flights covering both land and water surfaces versus flights over land surfaces only.

Figure 8a showcases a histogram of the total offline signal (in blue) and separate histograms for signals above land and water, respectively from a flight conducted on 12.06.2018. The entire flight covers land and ocean areas, as well as small segments over Alpine snow and the Sahara, as shown in the corresponding flight track in Figure 9a. The histograms of the offline signals above land show 3 distinct peaks with the smallest one near the lowest signal strength range which indicates Alpine regions in the map. The offline signal strength directly depends on the surface reflectance and higher reflectances lead to higher signal strengths. Therefore, over vegetation and forests with higher surface reflectance we get a higher offline signal strength indicated by the peak at 0.2 arbitrary units in Figure 8a. As the majority of the land signals originate from Italy's semi-arid soils and also

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

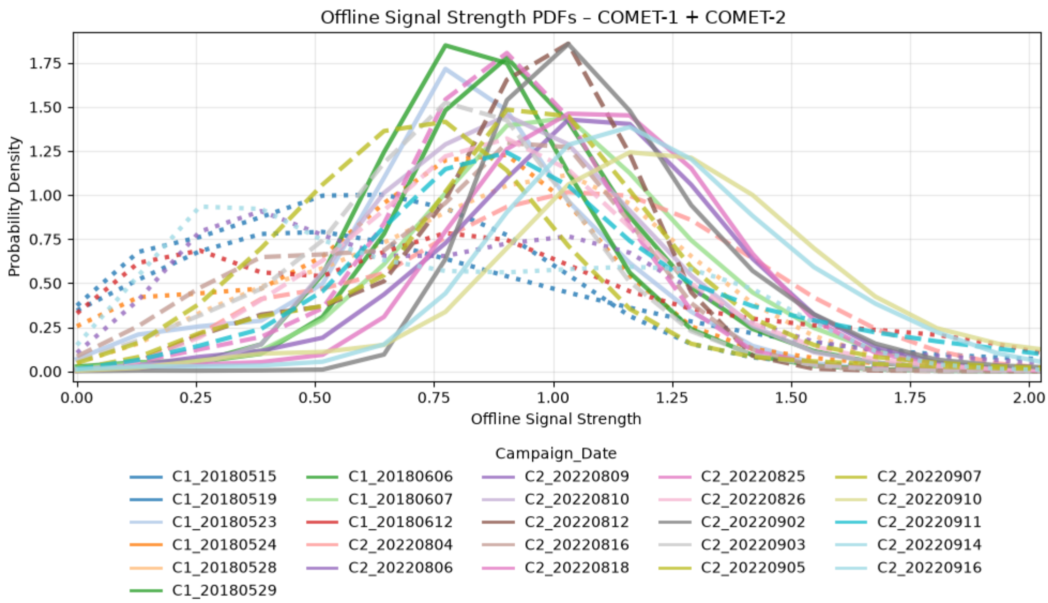
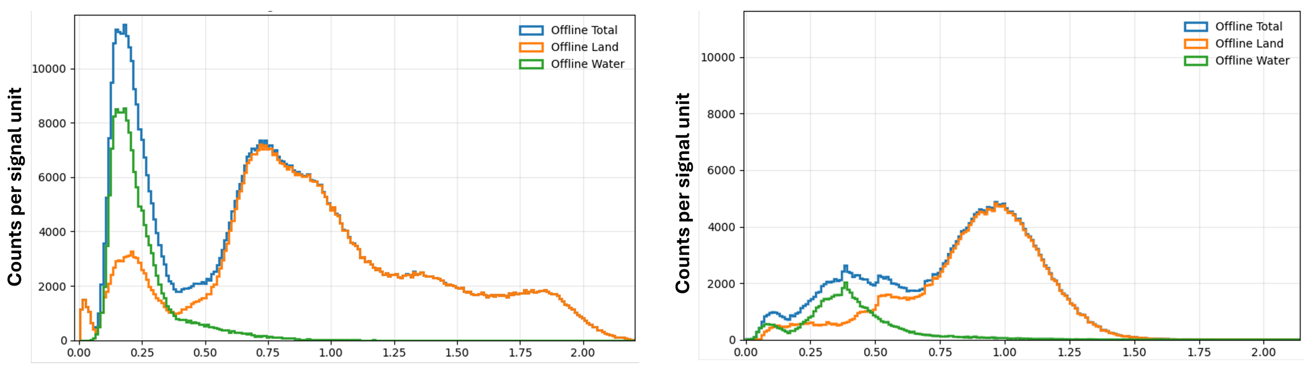


Figure 7. Probability density functions (PDFs) of the CHARM-F offline signal strength for all flights depicted in Figures 4 and 5. The abbreviations C1 and C2 denotes CoMet 1.0 and CoMet 2.0 campaigns, respectively. Each curve represents a single flight, distinguished by color and line style (solid for dry land, dashed for wetlands or inland water, and dotted for open ocean and sea). CoMet 1.0 flights over oceans and seas generally exhibit broader distributions with peaks at lower signal strengths, while CoMet 2.0 flights show more varied distributions sometimes with dual peaks for flights over wetlands. The spread across campaigns reflects differences in surface reflectance and atmospheric conditions.

deserts (in Northern Africa) , the PDF has a pronounced tail at high reflectances attributable to these areas. In contrast, the PDF of the signal strength above water surfaces in Figure 8a is mono-modal with a single maximum at low reflectivities between 0.1 and 0.3, also visible in the map in Figure 9a. The water and land PDFs combine to give an overall bimodal distribution. In conclusion, a large dynamic range is captured by CHARM-F flying over various surfaces, as expected.

Concerning the flight on 16.08.2022 during the CoMet 2.0 campaign, the histogram in Figure 8b and the map in Figure 9b show a similar variation of the offline signal over the entire research flight. We find a similar low reflectance region from the inland water bodies, giving a small peak in the histogram in the 0.2 - 0.5 signal strength range, as well as a dominant peak over high reflectance land surfaces. In total, the two dominant reflectance ranges again give a bimodal distribution of the offline signal strength for the whole flight, which leads to a non-Gaussian offline signal strength distribution. However, the non-Gaussian distribution does not affect the AVS scheme, since we weight with the corresponding offline and online SNR while averaging the measurements. The dynamic range is however smaller than in the CoMet 1.0 flight since there was no desert.

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026



(a) (b)

Figure 8. Histograms of CHARM-F Lidar Offline Signal Strength for one flight each from CoMet 1.0 over the central Mediterranean region (a) and CoMet 2.0 over western Canada (b), separated by surface type: total (blue), land (orange), and water (green).

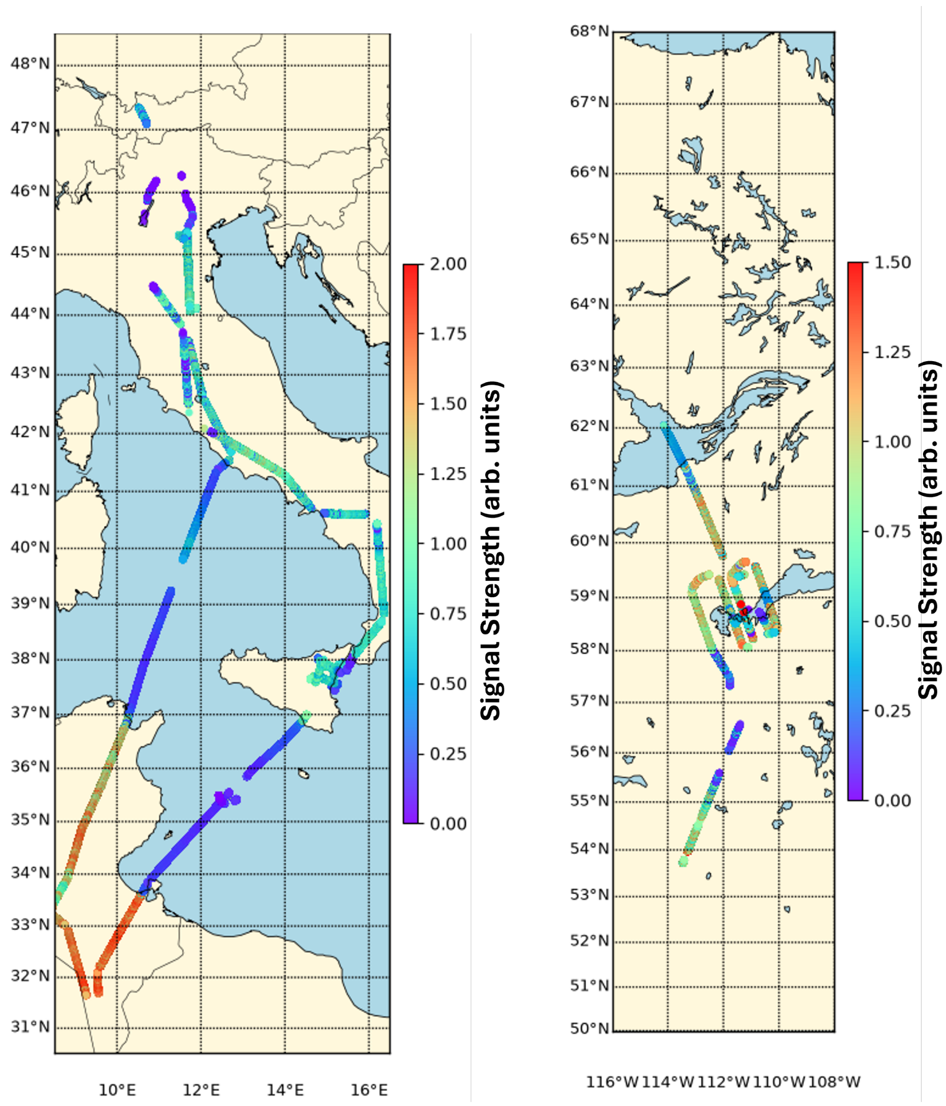
Overall, for the both scenarios we see a bimodal distribution with one peak at low signal strengths denoting water surfaces and one at a higher signal strength coming from vegetated surfaces or dry land. The dominance of the water peak in the total distribution in Figure 8a reflects the large fraction of over-water flight segments over the Mediterranean, whereas the long tail of the land distribution can be attributed to bright soil, desert and vegetation surfaces. The overall smaller and more compact distribution in Figure 8b compared to Figure 8a reflects more homogeneous land cover of the Canadian boreal survey region.

5.2 Comparison of specMACS surface reflectance vs. CHARM-F signals

In order to get to know more about the CHARM-F signals we compare the offline signal to the specMACS surface reflectances for a number of flights. The specMACS surface reflectance data has been processed for 5 flights, dated 12.08.2022, 16.08.2022, 18.08.2022, 26.08.2022 and 05.09.2022 from the CoMet 2.0 campaign. All flights cover both land and inland water surfaces, including wetlands in the Hudson Bay Lowlands during the flight on 05.09.2022 and Lake Winnipeg during the flight on 26.08.2022. Figure 10 shows an example of a timeseries of the specMACS surface reflectance against the CHARM-F offline signal strength for one of the CoMet 2.0 campaign flights on 16.08.2022. For the comparison with CHARM-F, specMACS measurements around the wavelength of $1645\,\mathrm{nm}$ were used, which is the closest wavelength to CHARM-F. To reduce the noise, the three closest specMACS wavelengths ($1645.66 \pm 1\,\mathrm{nm}$) have been averaged. Additionally, the nadir and the two closest to nadir pixels have been averaged, such that both footprints overlap. For further reducing the noise, a $3\,\mathrm{s}$ average on both data has been taken, which corresponds to 150 shots in CHARM-F and 90 frames in the specMACS measurements. The specMACS data are masked with the CHARM-F cloud mask to avoid cloud contamination to which the spectrometer is very sensitive.

Figure 10 shows a strong correlation between the specMACS surface reflectance (blue) and the CHARM-F offline signal strength (orange). Major drops in reflectance values corresponding to water bodies are matched by sharp decreases in the offline

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026



(a) (b)

Figure 9. Geographic map showing CHARM-F offline signal strength along the flight tracks for one flight each from CoMet 1.0 over the central Mediterranean region (a) and CoMet 2.0 over western Canada (b). The signal is color-coded, representing range normalized offline lidar signal strength. Note that the color scales are different for the two cases representing the whole range of signals. Higher signal values in both maps are predominantly observed over land surfaces, while lower values correspond to over-water flight segments over the Tyrrhenian, Adriatic, and Mediterranean Seas (a) or small inland water bodies, lakes and wetlands (b), consistent with the much lower NIR reflectance of water surfaces.

signal. The surface reflectance measured by specMACS reduces to values close to zero in these cases. The CHARM-F offline signal also shows reductions in the same time period (e.g., around 16:55 and 17:30), but the magnitude of this reduction is

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

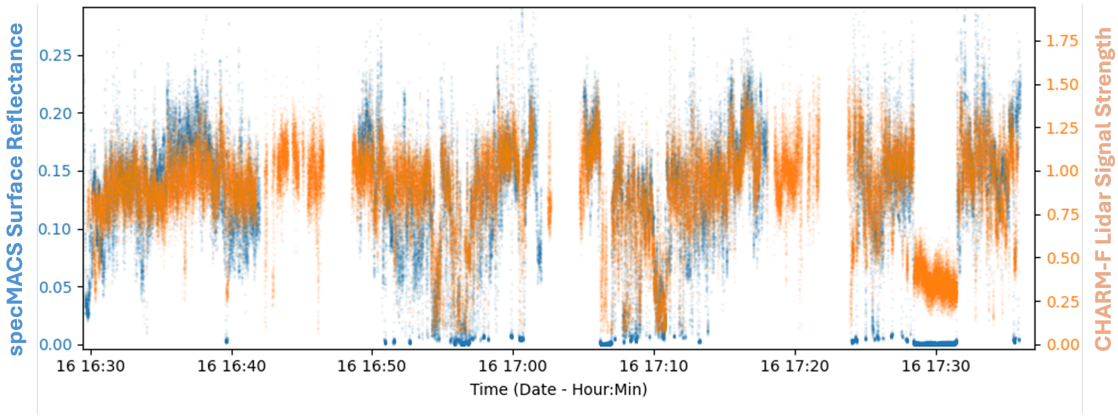


Figure 10. Time series comparison of specMACS surface reflectance averaged at 1645.66±1 nm (blue, left axis) and CHARM-F lidar offline signal strength (orange, right axis) for a representative cloud-free segment of the flight on 16.08.2022 from the CoMet 2.0 Arctic campaign.

smaller. The reason for this difference is that, the lidar is sensitive to specular and diffuse laser back scatter from the water surface and wave spray, whereas the reflectance measured by specMACS is strongly dependent on the solar incidence, dropping to values close to zero outside the glint angular range. The land surfaces generally maintain a surface reflectance between 0.10 and 0.25, which corresponds to an offline CHARM-F signal range of 0.75 to 1.50. Generally, the two sensors are not expected to agree identically in absolute terms. In fact, the explanation for the differences is that the bidirectional reflection distribution function (BRDF) or in general surface reflectance as called in this work has an angular dependence. It is close to zero except for a narrow angular range around the specular direction (known as sun glint). For the viewing geometry of specMACS, the sun glint is outside the field of view of the instrument and the water surfaces therefore appear very dark. In contrast to CHARM-F, the geometry of the emitted and received radiation over water corresponds to a partly specular reflection. Because the signal depends on the fraction of the illuminated area oriented perpendicular to the incoming beam, it manifests as laser glint (Shaw and Churnside (1997b), Shaw and Churnside (1997a)), resulting in a much higher signal over surfaces including water bodies (see also Fig. 2). But the general agreement between the two independent sensors validates the use of specMACS reflectance as a proxy for CHARM-F surface return variability.

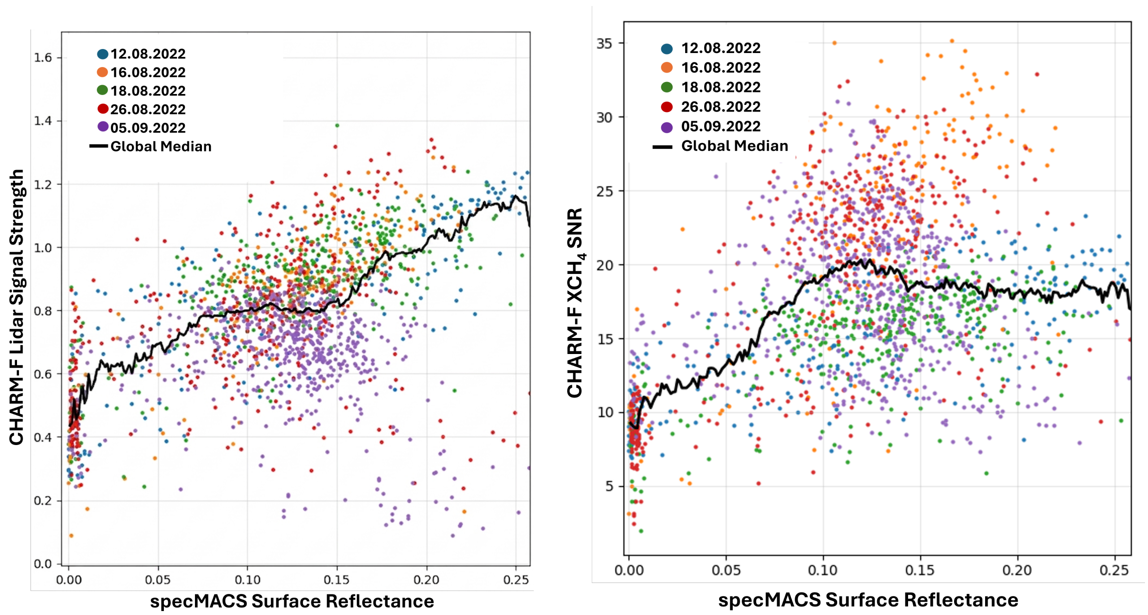
The scatter plot in Figure 11a illustrates the relationship between the specMACS surface reflectance and the CHARM-F range squared normalized offline signal with the global median line in black and a first degree polynomial line fitted to the global median curve for understanding the linearity of the data. Overall, the CHARM-F measurements exhibit a positive linear correlation with the specMACS surface reflectances which demonstrates that CHARM-F offline signal strength scales approximately linearly with surface reflectance. The y-intercept represents the offset between the specMACS surface reflectance and the CHARM-F offline signal strength above water surfaces due to their different geometries, as discussed above. In addition to this, we see two distinct clusters, one at a very low reflectance of ∼0.01-0.02 indicating water surfaces, and another at a higher reflectance range of 0.1 - 0.2 indicating vegetation covered land areas. The five different flights are denoted by different colors, which all covered different surface conditions and show different relationships between the specMACS surface reflectance and

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

the CHARM-F offline signal strength. A cluster of points from different flights can be seen above the global median line indicating stronger lidar signal compared to specMACS reflectance values. The flights on 16.08.2022, 18.08.2022 and 26.08.2022 cover bigger lakes and drier surfaces which leads to the cluster of points for these flights being more concentrated in a linear manner with the specMACS surface reflectance. On the other hand, there also exists a significant cluster of points below the global median which brings down the curve in the surface reflectance regime of 0.1 - 0.15. With the cluster of points of purple color lying majorly in the 0.6 - 0.8 offline signal strength range, it is evident that the flight on 05.09.2022 has lower offline signal strength even at higher reflectance regime as compared to the other flights. This may be attributed to the extended wetlands over which the flight took place which leads to mixed signals from vegetation and water under the vegetation. Additionally, a vegetation shadow effect on the wetland surfaces may also lower the lidar signal with respect to specMACS due to their different viewing geometries. We know that specMACS measures hemispherical directional reflectance factor for solar angle to nadir, whereas CHARM-F sees nadir to zenith. The increase in signal strength with surface reflectance for CHARM-F measurements is partly enhanced also by the hotspot effect, a bidirectional reflectance phenomenon that increases the back scattered lidar return when the illumination and viewing directions are nearly aligned (Amediek et al. (2009)). As for the sun glint and ocean surfaces, there can also be an angular dependence of the BRDF which indicates that the specMACS (its viewing geometry) as a proxy for CHARM-F might work better for some cases than for other.

Now we look into the variation of the CHARM-F XCH_4 SNR computed within the same window of 3 seconds using Eq. 9 as a function of specMACS surface reflectance as shown in Figure 11b with same colors as in Figure 11a. A positive dependence of SNR on surface reflectance is observed for very low reflectance values. At the near-zero specMACS reflectance around 0.01 - 0.02, we see a dense cluster of points from all flights and the global median SNR increases almost monotonically from $\sim$8 to $\sim$12. This regime shows the expected photon-limited behaviour of the IPDA lidar system in the sense that the returned number of photons increases with higher surface reflectance thereby improving the SNR. The spread is relatively large compared to the mean, indicating strong sensitivity to small variations in surface brightness. This regime likely corresponds to dark water bodies, wet soils, or shadowed surfaces where back scatter is minimal. For reflectance values of $\sim$0.10–0.16, the median SNR reaches a broad plateau ($\sim$17–25) and no longer increases substantially with reflectance. The SNR distribution also becomes more dispersed with few points even reaching a value of 35. This suggests a transition from a reflectance-driven (signal-limited) regime to a regime where other noise sources dominate (e.g., detector noise, speckle, atmospheric variability, or residual instrumental effects). Speckle noise here, mainly depends on the size of the illuminated area on the ground and scales with the size of the laser beam divergence (Cassé et al. (2019)). The flattening of the curve indicates that the SNR is not much affected by the increasingly bright surfaces. This regime likely corresponds to vegetated surfaces or mixed land–water pixels typical of the campaign region in Canada. We also see variability among the different flights. For example, the flights on 16.08.2022 and 26.08.2022 exhibit higher SNR values (frequent points above 20 and up to $\sim$35), suggesting either brighter surface conditions or more clear atmospheric transmission (less aerosol) during these flights. The flight on 05.09.2022 displays broad variability across all reflectance levels indicating the broad range of SNR signals over wetlands. However, most flights show SNR saturation above a certain reflectance threshold. In conclusion, the comparison reveals two distinct noise regimes with photon-noise-dominated behavior at low signal levels and a noise regime controlled by factors other than photon back

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026



(a) (b)

Figure 11. Scatter Plots of CHARM-F range^2 normalized offline signal strength (a) and XCH_4 signal-to-noise ratio (b) as a function of specMACS surface reflectance averaged over 3 seconds, for five research flights of the CoMet 2.0 campaign identified with different colours. The solid black line in both the plots represent the global median of all the data points. Additionally, the dashed black line on the left plot (a) indicates the first order linear fit to the median values of all the data points.

scatter at higher signal levels e.g. speckle noise. No major changes were undertaken from the instrument or laser point of view during the different flights in the campaign which can affect the variation in the SNR among the flights. Therefore, the spread in the SNR and the signal strength can be attributed majorly to the reasons explained above and not to any hardware changes.

5.3 Averaging Bias Characterization and Uncertainty Propagation for CHARM-F Signals

In this section the results from implementing the different averaging schemes formulated by Tellier et al. (2018) and described in Section 3 are compiled for flights from CoMet 1.0 and CoMet 2.0 campaigns. First, we do a similar comparison as Tellier et al. (2018) in estimating the initial bias using the different averaging schemes for all flights from the CoMet 2.0 campaign implementing the Equations 7 and 8. The averaging window for all the flights has been kept consistent at 3 seconds corresponding to 151 shots. From the analysis across all flights it is evident that the spatial averaging of signals approach in AVS is more robust under challenging retrieval conditions as can be seen from Figure A1. The mean value for initial estimated bias using AVS equals 0.8 ppb, whereas for AVX and AVD it stands at 2.3 ppb. This result emphasizes on the findings in the work of Tellier et al. (2018) showing that since the AVS scheme performs the averaging at the raw signal level before the non-linear logarithmic transformations, it is more robust to albedo variability across the averaging window. There is also a variation in

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

the magnitudes of the AVS bias between the different flights as described more in the Appendix section of Initial Bias Estimation. For some flights, we see very low biases in the range of 0.1-0.5 ppb while for others it can reach up to 2 ppb. This can be attributed to the surface heterogeneity or atmospheric variability during the flights. These initial estimated biases can be corrected for, while deriving the column averaged XCH_4. However, the uncertainty of the bias correction also depends on the flight scene and therefore, on the SNR and this is what we focus on from here. Since for airborne measurements as well, AVS comes out to be the best averaging method for lowest bias effect, which is in line with the results from the simulated measurements in Tellier et al. (2018), for further analysis and uncertainty propagation we consider only the AVS scheme.

We separate all the CoMet 1.0 and CoMet 2.0 flights into three categories, namely flights over dry land surfaces, over wetlands including inland lakes and over open water including sea. For further analysis, we compare the remaining uncertainty on the AVS XCH_4 estimate calculated from the uncertainty propagation as formulated in Section 3.5. The comparison is done to understand the effect of different scenes on the averaging bias and uncertainty estimation and also to understand the effect of separating the water signals from the dry land for averaging. For all analyses the averaging window is still kept at 3 sec (151 shots).

Figure 12 presents the mean non-correctable residual uncertainty against the XCH_4 SNR from all the flights of the two campaigns. The intermediate steps on the XCH_4 SNR estimates and uncertainties of XCH_4 estimates against the number of flights from the two campaigns are provided in the Appendix B. For total 11 flights over land from both campaigns, we can see that the aggregated SNR lies at 22.75 $\pm$ 3.13 resulting in an estimated, non-correctable residual uncertainty of 0.34 $\pm$ 0.32 ppb. The 8 flights over wetlands from the CoMet 2.0 campaign are processed in two different ways. In the first case, we use an inland water mask to separate the land signals from those over inland lakes and smaller inland water bodies which gives a lower SNR of 10.74 and a higher non-correctable residual uncertainty of 1.63 ppb. However, in the second case, where we do not separate the signals between dry land and wetlands, the SNR estimate is improved to 21.13 $\pm$ 2.84 and the uncertainty is reduced to 0.24 $\pm$ 0.11 ppb. The application of the inland water mask likely reflects the reduced number of valid retrieved shots remaining, where inland water bodies are spatially mixed with vegetated land. The reduced spatial sampling therefore, may affect the uncertainty statistics, increasing the uncertainty estimates. Consequently, in case of wetlands we propose to average all the measurements along the flight track including dry land and wetlands/inland water bodies to achieve a better SNR and a smaller residual uncertainty similar to those over the land alone flights.

The flights over open extended water (North Sea and Mediterranean Sea) are processed after masking with the water masks so we only deal with measurements over water. However, in this case when using the external water mask and process only the shots over extended water body separately, we get a lower mean uncertainty of 0.45 ppb as compared to an uncertainty of 1.68 ppb if the measurements over land and water are averaged together. This in turn results in a higher SNR value of 14.48 $\pm$ 2.76. The application of a water mask in this case, therefore, reduces the uncertainties and demonstrates that for longer flight paths over extended water bodies it is required to handle the processing of measurements over water separately due to a different range of signal strength and SNR over water as compared to dry land. Based on the uncertainty propagation formulation, it can be said that if the SNR goes to extreme values inside the averaging window because of changing scenes from land to water (coastlines) or changing sea surface roughness based on surface winds, it is more probable to get a higher

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

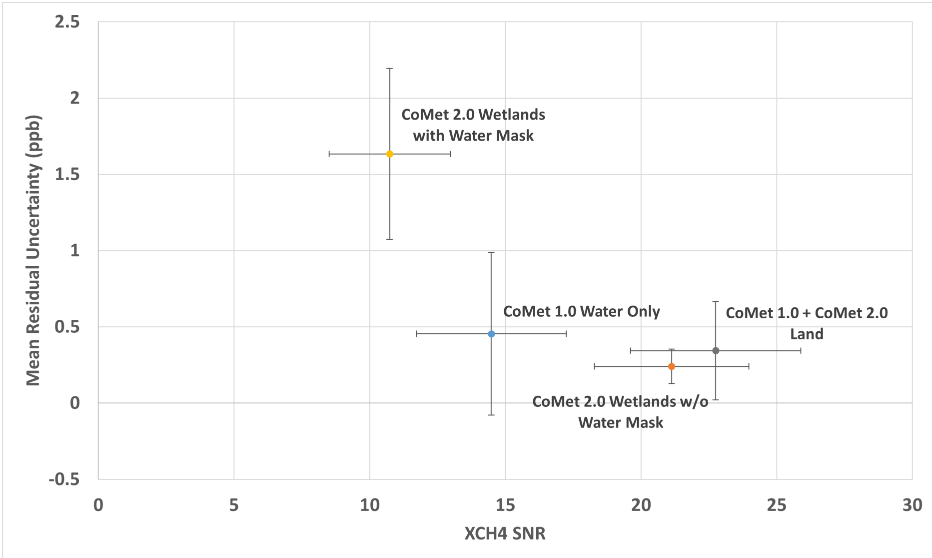


Figure 12. Scatter plot of XCH_4 SNR vs. mean non-correctable residual uncertainty (ppb) for four CoMet retrieval configurations identified with different colours. Error bars represent the standard deviation in both dimensions. It clearly indicates that higher signal quality is associated with reduced retrieval uncertainty.

Table 2. Summary of Initial Bias Estimation with 3 Averaging Schemes and Final Non-correctable Residual Uncertainty for AVS for CoMet 2.0 flights

Scene	AVX/AVD Initial Bias	AVS Initial Bias	AVS Residual Uncertainty
Dry Land	0.7 ppb ($\sim$ 0.03%)	0.3 ppb ($\sim$ 0.02%)	0.2 ppb ($\sim$ 0.01%)
Wetlands/Water	2.7 ppb ($\sim$ 0.14%)	0.8 ppb ($\sim$ 0.04%)	0.3 ppb ($\sim$ 0.02%)

525 uncertainty on the XCH_4 estimate. Therefore, the measurements over ocean or open water bodies have to be dealt differently from wetlands.

Taken together, the CoMet 1.0 results reinforce the conclusion drawn from CoMet 2.0, showing that the presence of open water or specular-reflecting surfaces in the flight tracks can be a primary driver of lower SNR and slightly higher residual uncertainty in AVS-derived XCH_4 estimates, irrespective of specific flight conditions. This finding highlights the need of 530 applying a water mask over open ocean (i.e. extended water bodies) as compared to complex wetland environments, pointing out the need for scene-specific masking strategies in future retrieval algorithms.

The complete picture of the initial bias estimation for the three averaging schemes for CoMet 2.0 flights and the final non-correctable uncertainty values is given in Table 2 considering a background XCH_4 of 1900 ppb.

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

6 Conclusions

This study presents the first observationally constrained evaluation of averaging-induced biases and uncertainty propagation in CHARM-F IPDA lidar XCH_4 retrievals across a comprehensive range of surface types, using measurements from the CoMet 1.0 and CoMet 2.0 airborne campaigns. By combining CHARM-F lidar measurements with independent specMACS hyperspectral surface reflectance observations and auxiliary water mask datasets, we characterize the relation between surface type, signal statistics, averaging scheme performance, and retrieval uncertainty.

The CHARM-F offline signal strength scales linearly with specMACS surface reflectance, with a positive intercept (∼0.55 a.u.) which can be attributed to significantly higher lidar back scatter from low-albedo surfaces that are optically dark to passive sensors. The back scattered signal from the reflecting water surface has a strong angular dependence for passive sensor specMACS which gives differences in the reflectance values inside and outside the sun glint angular range. This demonstrates CHARM-F's unique capability to retrieve methane signals over water surfaces challenging passive instruments.

At low reflectances (<0.05), the lidar signal SNR is photon-noise-limited and increases steeply with surface brightness, making retrieval quality highly sensitive to surface albedo variability. At moderate to high reflectances (>0.10), the SNR saturates and is controlled by other noise sources such as detector or speckle noise. Retrieval strategies must account for both regimes, particularly for wetland and coastal environments where both regimes may occur within a single averaging window.

The study focuses on averaging biases and uncertainties in XCH_4 estimates arising from the non-linearity of the DAOD retrieval under measurement noise and how it is correlated to the XCH_4 SNR. We exclude the detector non-linearity and saturation effects from the scope of the analysis as also mentioned in Section 5.3. The signal-level averaging approach of AVS, which accumulates online and offline signals before applying the nonlinear DAOD transformation, consistently produces the lowest averaging bias for all flight conditions investigated, in agreement with simulation-based deductions of Tellier et al. (2018). This advantage is most pronounced over low-reflectance surfaces, where AVX and AVD can produce mean biases in the range of 2-3 ppb ($\sim 0.11\%$ - 0.16% considering a background XCH_4 of 1900 ppb) due to their sensitivity to noisy signals at the single-shot level. AVS, therefore, proves to be the best averaging scheme for both airborne CHARM-F and future MERLIN space-borne retrievals with a mean initial bias estimation of 0.8 ppb ($\sim 0.04\%$).

After the correction of the estimated bias with the AVS scheme, the remaining uncertainty on the XCH_4 estimate then depends on the SNR variability among the different surfaces. Dry land flights from both campaigns yield non-correctable residual XCH_4 uncertainties well below 0.7 ppb ($< 0.03\%$) with AVS, consistent with the high SNR (17-28) characteristics of vegetated and semi-arid land surfaces. Wetland and ocean flights exhibit slightly higher uncertainties but still below 1 ppb ($< 0.05\%$), reflecting SNR values of 8-15 and the bimodal, non-Gaussian offline signal distributions associated with mixed land–water scenes. This shows that surface type impacts the retrieval uncertainty.

Over extended open water bodies, applying a water mask and processing measurements separately reduces the XCH_4 retrieval uncertainty by removing the heterogeneous influence of low-SNR shots from land-dominated averaging windows. In complex wetland environments, however, the inland water masking reduces sample counts to the point, where SNR estimation becomes unstable and uncertainty estimates increase. In these environments, combining land and water measurements within

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

the averaging window rather than segregating them yields retrievals with lower uncertainty. The land–water–wetland contrasts and this surface-dependent masking behavior must be explicitly accounted for in calibration and data processing schemes for future MERLIN retrieval algorithm design.

These findings, therefore, not only provide guidance in CHARM-F dataset interpretation but also valuable lessons for space borne active remote sensing missions like the upcoming MERLIN mission. The combination of robust averaging schemes with SNR adaptive processing will provide space-borne IPDA retrievals with minimum uncertainty across all Earth's surface types.

7 Outlook

The results of this study open up several directions for further investigation, both in the context of continued studies on CHARM-F measurements and in preparation for the MERLIN satellite mission.

The current analysis is necessarily limited by the number of available flights with over-water or over-wetland segments. Additionally, surface types with very low SNR such as sea ice and snow-covered terrain, which were explicitly excluded from the current analysis, also represent important test cases for extreme-low-reflectance retrieval performance and should be characterized in future work on the basis of more dedicated flights.

The finding that the uncertainty estimates depend on the SNR from different surface types motivates the development of SNR-adaptive retrieval frameworks in which the averaging window size is selected dynamically based on estimates of surface reflectance and signal quality. A hybrid scheme that applies AVS globally but adjusts the effective averaging window locally in response to measured SNR, for example, widening the window over low-reflectance surfaces to accumulate sufficient photons, offers a promising path to improving uncertainties in XCH_4 estimates.

Future work can also combine the averaging bias correction framework developed here with high resolution land surface classification data to separate signals over land into different types of vegetation cover, and relate wind-driven surface roughness models for water bodies to correlate the residual XCH_4 uncertainty with different wave height or surface wind patterns. Finally, a simulation study to translate the findings from the CHARM-F measurements to MERLIN observations and understand the detection limit of MERLIN over water would be the right direction to follow establishing a complete picture of detection limits of methane emissions for MERLIN satellite over different surface types.

Data availability. The CHARM-F observations from the two CoMet campaigns and the specMACS imaging dataset can be directly obtained from the authors. The publication has been done using water mask datasets from Copernicus Land Management System (https://doi.org/10.2909/86c14646-c0b6-4b82-a82f-cbb23b331743, https://doi.org/10.2909/4cd93293-e944-4046-987c-66e4f59a2071, last access: 12 March, 2026).

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

Appendix A: Initial Bias Estimation

This details the initial bias estimation using the 3 different averaging schemes for all flights from the CoMet 2.0 campaign. Figure A1 shows a comparative bar chart on the averaging bias resulting from the 3 different averaging schemes in 3 colors: blue (AVS), orange (AVX) and gray (AVD). It is evident that AVX and AVD result in higher biases as compared to the AVS
600 method for all flights.

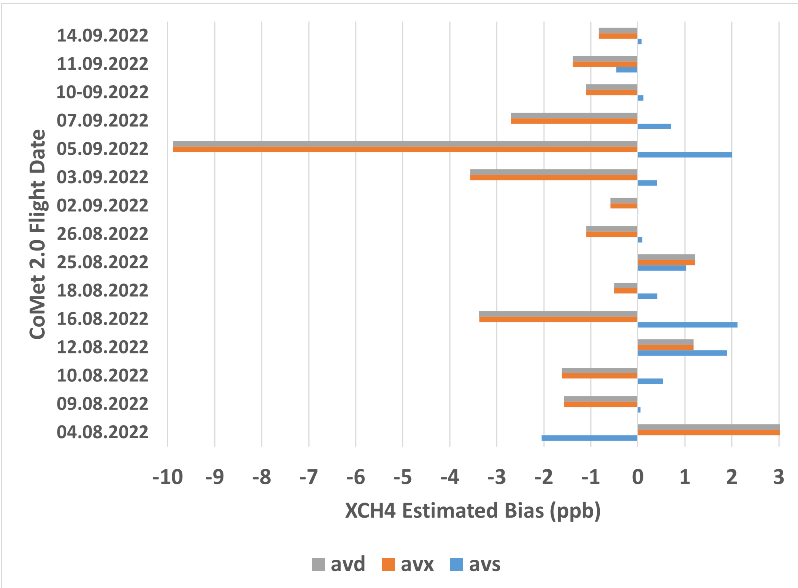


Figure A1. Comparison of Averaging Bias in ppb for all flights from the CoMet 2.0 campaign across the 3 different averaging schemes (AVS, AVD, AVX). The average initial bias for AVS comes to 0.8 ppb whereas for AVX and AVD it comes to 2.25 ppb.

Appendix B: Residual Uncertainty Characterization

Figure B1 presents the XCH_4 retrieval uncertainty for dry land surface flights conducted during both the CoMet 1.0 (2018) and CoMet 2.0 (2022) campaigns. CoMet 1.0 flights (shown in orange) exhibit slightly higher uncertainties, ranging from approximately 0.09 ppb (07.06.2018) to a maximum of ∼1.17 ppb (06.06.2018). This large spread in CoMet 1.0 uncertainties
605 suggests notable variability in retrieval biases across those flights, potentially driven by differences in surface reflectance, atmospheric conditions, or scene heterogeneity over land from desert to vegetation and minute snow cover. In contrast, CoMet 2.0 flights (shown in blue) demonstrate substantially lower and more consistent uncertainties, all falling below 0.4 ppb. The observed improvement of the mean residual non-correctable uncertainty above land from CoMet 1.0 to CoMet 2.0 can be attributed to advances in the instrument, minor changes to the retrieval algorithm approach and different atmospheric conditions.
610 The majority cluster between ∼0.10 and ∼0.18 ppb (flights from 04.08.2022 through 14.09.2022), with the notable exception

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

of 25.08.2022, which shows an uncertainty of $\sim$0.37 ppb, and 02.09.2022, which shows the lowest value at $\sim$0.05 ppb. The overall range for both campaigns stays within 1 ppb showing a much lower bias for XCH_4 estimate over dry land.

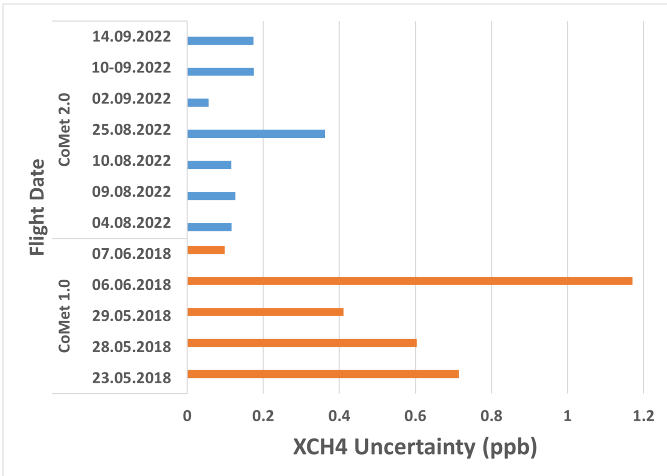


Figure B1. Residual Uncertainty on the XCH_4 estimate for all flights over dry land from CoMet 1.0 (orange) and CoMet 2.0 campaigns (blue)

Figure B2 shows XCH_4 retrieval uncertainties for CoMet 2.0 wetland flights, comparing results obtained with an inland water mask (blue) against those without (orange), across eight flight dates from 12.08.2022 to 11.09.2022. With the inland water mask applied (blue bars), uncertainties are slightly higher across all flights indicating the same range as the original bias values. The application of the inland water mask likely reflects the reduced number of valid retrieved shots remaining, where inland water bodies are spatially interspersed with vegetated land. The reduced spatial sampling therefore, may destabilize the retrieval statistics, increasing the uncertainty estimates. Without the inland water mask (orange bars), uncertainties are uniformly low, ranging from approximately 0.07 ppb (03.09.2022 and 11.09.2022) to $\sim$0.45 ppb (16.08.2022), with most flights falling between 0.15 and 0.35 ppb. The idea in case of wetlands is to, therefore, average all the measurements along the flight track including dry land and wetlands/inland water bodies to achieve the smaller uncertainties similar to those over land alone flights. This suggests that without masking, the wetland retrievals are relatively well-constrained when inland water pixels are not treated separately but rather combined with neighboring signals from dry land.

Figure B3 illustrates the impact of applying a water mask on XCH_4 retrieval uncertainty for sea and ocean surface CoMet 1.0 flights on three dates: 15.05.2018, 19.05.2018, and 12.06.2018. Results are grouped into two categories namely, uncertainty calculation performed without a water mask (blue) and with a water mask applied (orange). Without the water mask, uncertainties are slightly higher, reaching $\sim$2.85 ppb on 15.05.2018 where we had a considerable section of the flight above North Sea. The 12.06.2018 flight having a wide range of signal strength (as shown before in figure 8a) shows a moderately low uncertainty value when estimating the XCH_4 over the entire flight track. These values are consistent with the expected difficulty of retrieving XCH_4 over specularly reflecting or low-albedo ocean surfaces, where the signal-to-noise ratio is reduced.

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

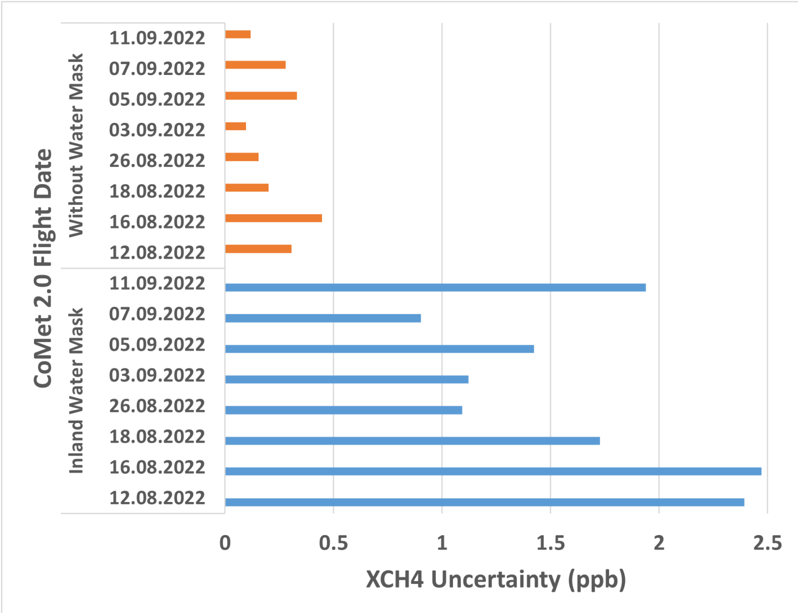


Figure B2. Residual Uncertainty on the XCH_4 estimate for all flights over wetlands and lakes from CoMet 2.0 campaign. The bar charts in blue indicate the flights have been masked with inland water mask and the uncertainty has been calculated on measurements only over the inland water points. The bar charts in orange indicate the flights have not been masked with inland water mask and the uncertainty has been calculated on measurements combining dry land and inland water points.

The application of a water mask in this case actually reduces the uncertainties across all three flights. This demonstrates that for longer flight paths over extended water bodies, handling the processing of measurements over water separately is critical. Therefore, the case for ocean or open water bodies has to be dealt differently from wetlands since the ocean or sea surface is also comparatively more rough than the inland lakes from the effect of the wind at the water surface.

Figure B4 represents the mean XCH_4 SNR for four CoMet retrieval configurations across varying numbers of validation flights. The CoMet 1.0 + CoMet 2.0 Land configuration achieved the highest mean XCH_4 SNR (22.75), followed closely by CoMet 2.0 Wetlands without Mask (21.13). CoMet 1.0 Water Only yielded an intermediate SNR (14.48). The CoMet 2.0 Wetlands with Mask configuration recorded the lowest SNR (10.74), consistent with its elevated retrieval uncertainty.

Figure B5 presents the mean non-correctable residual uncertainty against the number of flights used for each retrieval configuration from the two campaigns. The CoMet 1.0 + CoMet 2.0 Land configuration achieved the lowest mean residual uncertainty (0.34 ppb, n=11 flights), with relatively narrow spread across flights. CoMet 2.0 Wetlands without Mask and CoMet 1.0 Open Water with Mask exhibited comparable intermediate uncertainty levels (0.24 ppb and 0.45 ppb, respectively), though the open water configuration was evaluated over fewer flights (n=3). In contrast, the CoMet 2.0 Wetlands with Mask configuration showed higher mean residual uncertainty (1.63 ppb), along with substantially wider error bars, indicating greater flight-to-flight variability and reduced retrieval consistency over masked wetland surfaces. The different variability in the final non-correctable

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

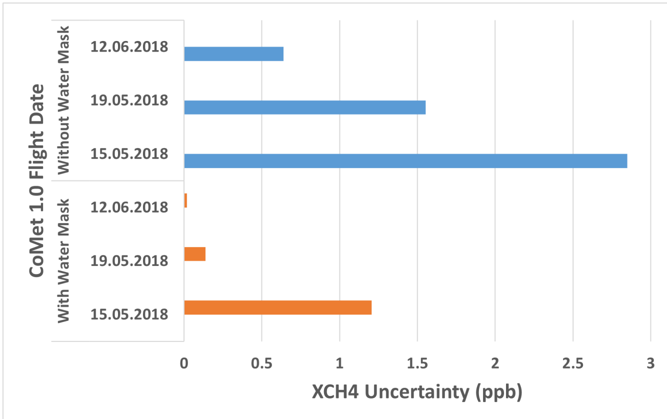


Figure B3. Residual Uncertainty on the XCH_4 estimate for all flights over longer stretches of water (North Sea and Mediterranean Sea) from CoMet 1.0 campaign. The bar charts in orange indicate the flights have been masked with external water mask and the uncertainty has been calculated on measurements only over the extensive stretches of water bodies. The bar charts in blue indicate the flights have not been masked with external water mask and the uncertainty has been calculated on measurements combining dry land and water points.

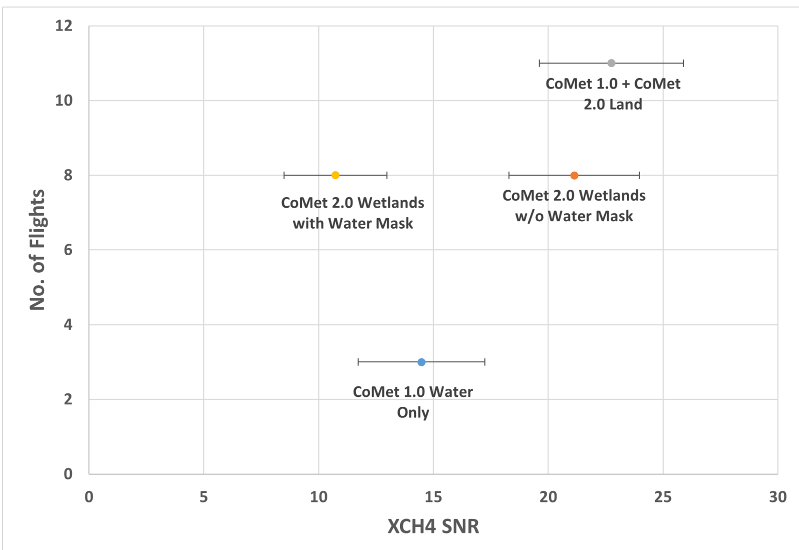


Figure B4. Mean XCH_4 SNR for four CoMet retrieval configurations versus the number of flights. Error bars represent the standard deviation across flights. Configurations shown are: CoMet 1.0 + CoMet 2.0 Land (grey, n=11), CoMet 2.0 Wetlands without Water Mask (orange, n=8), CoMet 2.0 Wetlands with Water Mask (yellow, n=8), and CoMet 1.0 Water Only (blue, n=3).

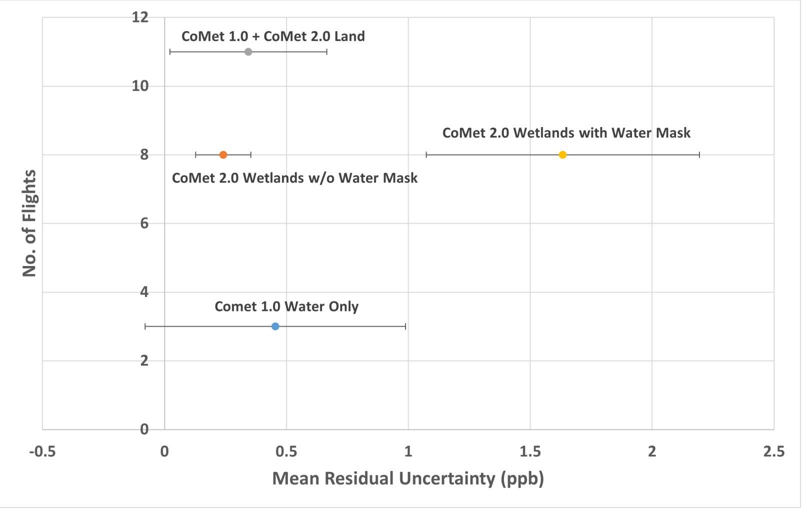


Figure B5. Mean residual uncertainty (ppb) for four CoMet retrieval configurations versus the number of flights. Error bars represent the standard deviation across flights. Configurations shown are: CoMet 1.0 + CoMet 2.0 Land (grey, n=11), CoMet 2.0 Wetlands without Water Mask (orange, n=8), CoMet 2.0 Wetlands with Water Mask (yellow, n=8), and CoMet 1.0 Open Water with Water Mask (blue, n=3).

residual uncertainty across the different scenarios can be potentially driven by differences in surface reflectance, atmospheric conditions, and scene heterogeneity.

Author contributions. The planning and execution of the CoMet measurement campaigns were carried out by A.F., M.Q., C.K. and S.Z. A.D. performed the formal data analysis and derived the final results. A.D., A.W. and C.K. contributed to the validation and interpretation of the results. The original draft of the manuscript was prepared by A.D. and critically reviewed and edited by all co-authors.

Competing interests. There are no competing interests present.

Acknowledgements. This work was funded by the Federal Ministry of Research, Technology and Space of Germany under the funding no. FKZ: 50EE2420. The campaigns CoMet 1.0 and CoMet 2.0 Arctic were supported by the priority program SPP 1294, High Altitude and Long-Range Research Aircraft (HALO) which is funded by the German Research Foundation (Deutsche Forschungsgemeinschaft - DFG). The atmospheric correction of the specMACS data was performed by Czech Globe within the ESA project CAIRTEX. In addition, we acknowledge funding by the Deutsche Forschungsgemeinschaft within the project SPP 1294, grant no. 442667104. This publication has been prepared using European Union's Copernicus Land Monitoring Service information (Pekel et al. (2016), CLMS-Water-Bodies (2022)).

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

References

Amediek, A., Fix, A., Ehret, G., Caron, J., and Durand, Y.: Airborne lidar reflectance measurements at 1.57 μm in support of the A-SCOPE mission for atmospheric CO_2, Atmospheric Measurement Techniques, 2, 755–772, https://doi.org/10.5194/amt-2-755-2009, 2009.

Amediek, A., Ehret, G., Fix, A., Wirth, M., Büdenbender, C., Quatrevalet, M., Kiemle, C., and Gerbig, C.: CHARM-F: A new airborne integrated-path differential-absorption lidar for carbon dioxide and methane observations: measurement performance and quantification of strong point source emissions, Appl. Opt., 56, 5182–5197, https://doi.org/10.1364/AO.56.005182, 2017.

Barkley, Z., Davis, K., Miles, N., Richardson, S., Deng, A., Hmiel, B., Lyon, D., and Lauvaux, T.: Quantification of oil and gas methane emissions in the Delaware and Marcellus basins using a network of continuous tower-based measurements, Atmospheric Chemistry and Physics, 23, 6127–6144, https://doi.org/10.5194/acp-23-6127-2023, 2023.

Bartsch, A., Gay, B. A., Schüttemeyer, D., Malina, E., Miner, K., Grosse, G., Fix, A., Tamminen, J., Bösch, H., Parker, R. J., Rautiainen, K., Hashemi, J., and Miller, C. E.: Advancing the Arctic Methane Permafrost Challenge (AMPAC) With Future Satellite Missions, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 6279–6298, https://doi.org/10.1109/JSTARS.2025.3538897, 2025.

Boy, M., Zhou, P., Kurtén, T., Chen, D., Xavier, C., Clusius, P., Roldin, P., Baykara, M., Pichelstorfer, L., Foreback, B., Bäck, J., Petäjä, T., Makkonen, R., Kerminen, V.-M., Pihlatie, M., Aalto, J., and Kulmala, M.: Positive feedback mechanism between biogenic volatile organic compounds and the methane lifetime in future climates, npj Climate and Atmospheric Science, 5, https://doi.org/10.1038/s41612-022-00292-0, 2022.

Bréon, F. M. and Henriot, N.: Spaceborne observations of ocean glint reflectance and modeling of wave slope distributions, Journal of Geophysical Research: Oceans, 111, https://doi.org/https://doi.org/10.1029/2005JC003343, 2006.

Cassé, V., Gibert, F., Edouart, D., Chomette, O., and Crevoisier, C.: Optical Energy Variability Induced by Speckle: The Cases of MERLIN and CHARM-F IPDA Lidar, Atmosphere, 10, 540, https://doi.org/10.3390/atmos10090540, 2019.

CLMS-Water-Bodies: Copernicus Global Land Monitoring Service Water Bodies 100m v1 monthly, https://doi.org/10.2909/86c14646-c0b6-4b82-a82f-cbb23b331743, 2022.

Copernicus Data Space Ecosystem: https://doi.org/10.5270/ESA-c5d3d65, © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018 provided under COPERNICUS by the European Union and ESA; all rights reserved, 2021.

Ehret, G., Kiemle, C., Wirth, M., Amediek, A., Fix, A., and Houweling, S.: Space-borne remote sensing of CO2, CH4, and N2O by integrated path differential absorption lidar: A sensitivity analysis, Applied Physics B, 90, 593–608, https://doi.org/10.1007/s00340-007-2892-3, 2008.

Ehret, G., Bousquet, P., Pierangelo, C., Alpers, M., Millet, B., Abshire, J., Bovensmann, H., Burrows, J., Chevallier, F., Ciais, P., Crevoisier, C., Fix, A., Flamant, P. H., Frankenberg, C., Gibert, F., Heim, B., Heimann, M., Houweling, S., Hubberten, H., Jöckel, P., Law, K. S., Löw, A., Marshall, J., Agusti-Panareda, A., Payan, S., Prigent, C., Rairoux, P., Sachs, T., Scholze, M., and Wirth, M.: MERLIN: A French-German Space Lidar Mission Dedicated to Atmospheric Methane, Remote Sensing, 9, 1052, https://doi.org/10.3390/rs9101052, 2017.

Ewald, F., Kölling, T., Baumgartner, A., Zinner, T., and Mayer, B.: Design and characterization of specMACS, a multipurpose hyperspectral cloud and sky imager, Atmospheric Measurement Techniques, 9, 2015–2042, https://doi.org/10.5194/amt-9-2015-2016, 2016.

Fiehn, A., Eckl, M., Pühl, M., Bräuer, T., Gottschaldt, K.-D., Aufmhoff, H., Eirenschmalz, L., Neumann, G., Sakellariou, F., Sauer, D., Baumann, R., De Aguiar Ventura, G., Cadete, W. N., Zua, D. L., Xavier, M., Correia, P., and Roiger, A.: Airborne quantification of

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

Angolan offshore oil and gas methane emissions, Atmospheric Chemistry and Physics, 25, 15 009–15 031, https://doi.org/10.5194/acp-25-15009-2025, 2025.

Fix, A., Büdenbender, C., Wirth, M., Quatrevalet, M., Amediek, A., Kiemle, C., and Ehret, G.: Optical Parametric Oscillators and Amplifiers for Airborne and Spaceborne Active Remote Sensing of CO2 and CH4, vol. 8182, p. 818206, ISBN 9780819488091, https://doi.org/10.1117/12.898412, 2011.

Fix, A., Quatrevalet, M., Amediek, A., and Wirth, M.: Energy calibration of integrated path differential absorption lidars, Appl. Opt., 57, 7501–7514, https://doi.org/10.1364/AO.57.007501, 2018.

Grinham, A., Albert, S., Deering, N., Dunbabin, M., Bastviken, D., Sherman, B., Lovelock, C. E., and Evans, C. D.: The importance of small artificial water bodies as sources of methane emissions in Queensland, Australia, Hydrology and Earth System Sciences, 22, 5281–5298, https://doi.org/10.5194/hess-22-5281-2018, 2018.

HALO CoMet 2.0 Arctic campaign: https://comet2arctic.de/, last access: 12 February 2026.

HALO CoMet campaign: https://www.halo.dlr.de/science/missions/comet/comet.html, last access: 12 February 2026.

Kort, E., Wofsy, S., Daube, B., Diao, M., Elkins, J., Gao, R., Hintsa, E., Hurst, D., Jimenez, R., Moore, F., Spackman, J., and Zondlo, M.: Atmospheric observations of Arctic Ocean methane emissions up to 82° north, Nature Geoscience, 5, 318, https://doi.org/10.1038/ngeo1452, 2012.

Krautwurst, S., Fruck, C., Wolff, S., Borchardt, J., Huhs, O., Gerilowski, K., Gałkowski, M., Kiemle, C., Quatrevalet, M., Wirth, M., Mallaun, C., Burrows, J. P., Gerbig, C., Fix, A., Bösch, H., and Bovensmann, H.: Identification and quantification of CH_4 emissions from Madrid landfills using airborne imaging spectrometry and greenhouse gas lidar, Atmospheric Chemistry and Physics, 25, 14 669–14 702, https://doi.org/10.5194/acp-25-14669-2025, 2025.

Li, M., Jackson, R. B., Saunois, M., Ciais, P., Poulter, B., Canadell, J. G., Patra, P. K., Tian, H., Zhang, Z., Fluet-Chouinard, E., Ouyang, Z., Zhang, T., Beerling, D. J., Belikov, D. A., Bousquet, P., Custodio, D., Chandra, N., Dou, X., Gedney, N., Hopcroft, P. O., Hoyt, A. M., Ichii, K., Ito, A., Jain, A. K., Jensen, K., Joos, F., Kleinen, T., Kondo, M., Li, F., Li, T., Liu, X., Maksyutov, S., Malhotra, A., Martinez, A., McDonald, K., Melton, J. R., Müller, J., Niwa, Y., Pan, S., Peng, S., Peng, C., Qin, Z., Raymond, P., Riley, W., Segers, A., Thompson, R. L., Tsuruta, A., Xi, Y., Yuan, K., Zhang, W., Zheng, B., Zhu, Q., Zhu, Q., and Zhuang, Q.: Machine-learning-based estimates of global natural vegetated wetland methane emissions (2000–2025), Earth System Science Data, 18, 3507–3524, https://doi.org/10.5194/essd-18-3507-2026, 2026.

Lin, X., Peng, S., Ciais, P., Hauglustaine, D., Lan, X., Liu, G., Ramonet, M., Xi, Y., Yin, Y., and Zhang, Z.: Recent methane surges reveal heightened emissions from tropical inundated areas, Nature Communications, 15, 2041–1723, https://doi.org/10.1038/s41467-024-55266-y, 2024.

Lorente, A., Borsdorff, T., Martinez-Velarte, M. C., Butz, A., Hasekamp, O. P., Wu, L., and Landgraf, J.: Evaluation of the methane full-physics retrieval applied to TROPOMI ocean sun glint measurements, Atmospheric Measurement Techniques, 15, 6585–6603, https://doi.org/10.5194/amt-15-6585-2022, 2022.

MacLean, J.-P. W., Girard, M., Jervis, D., Marshall, D., McKeever, J., Ramier, A., Strupler, M., Tarrant, E., and Young, D.: Offshore methane detection and quantification from space using sun glint measurements with the GHGSat constellation, Atmospheric Measurement Techniques, 17, 863–874, https://doi.org/10.5194/amt-17-863-2024, 2024.

McNicol, G., Fluet-Chouinard, E., Ouyang, Z., Knox, S., Zhang, Z., Aalto, T., Bansal, S., Chang, K.-Y., Chen, M., Delwiche, K., Feron, S., Goeckede, M., Liu, J., Malhotra, A., Melton, J. R., Riley, W., Vargas, R., Yuan, K., Ying, Q., Zhu, Q., Alekseychik, P., Aurela, M., Billesbach, D. P., Campbell, D. I., Chen, J., Chu, H., Desai, A. R., Euskirchen, E., Goodrich, J., Griffis, T., Helbig, M., Hirano, T., Iwata,

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

H., Jurasinski, G., King, J., Koebsch, F., Kolka, R., Krauss, K., Lohila, A., Mammarella, I., Nilson, M., Noormets, A., Oechel, W., Peichl, M., Sachs, T., Sakabe, A., Schulze, C., Sonnentag, O., Sullivan, R. C., Tuittila, E.-S., Ueyama, M., Vesala, T., Ward, E., Wille, C., Wong, G. X., Zona, D., Windham-Myers, L., Poulter, B., and Jackson, R. B.: Upscaling Wetland Methane Emissions From the FLUXNET-CH4 Eddy Covariance Network (UpCH4 v1.0): Model Development, Network Assessment, and Budget Comparison, AGU Advances, 4, e2023AV000 956, https://doi.org/https://doi.org/10.1029/2023AV000956, e2023AV000956 2023AV000956, 2023.

MERLIN Satellite Mission: https://merlin-methane.space/mission/satellite, last access: 29 May 2026.

Mwanake, R. M., Gettel, G. M., Wangari, E. G., Glaser, C., Houska, T., Breuer, L., Butterbach-Bahl, K., and Kiese, R.: Anthropogenic activities significantly increase annual greenhouse gas (GHG) fluxes from temperate headwater streams in Germany, Biogeosciences, 20, 3395–3422, https://doi.org/10.5194/bg-20-3395-2023, 2023.

Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A.: High-resolution mapping of global surface water and its long-term changes, Nature, 540, 418–422, https://doi.org/10.1038/nature20584, 2016.

Richter, R.: A spatially adaptive fast atmospheric correction algorithm, International Journal of Remote Sensing, 17, 1201–1214, https://doi.org/10.1080/01431169608949077, 1996.

Richter, R. and Schläpfer, D.: Atmospheric / Topographic Correction for Airborne Imagery, http://www.rese.ch/pdf/atcor4_manual.pdf, dLR-IB 565-02/15, 2016.

Richter, R., Schläpfer, D., and Müller, A.: An automatic atmospheric correction algorithm for visible/NIR imagery, International Journal of Remote Sensing, 27, 2077–2085, https://doi.org/10.1080/01431160500486690, 2006.

Riddick, S. N., Mbua, M., Laughery, C., and Zimmerle, D. J.: A Review of Offshore Methane Quantification Methodologies, Atmosphere, 16, https://doi.org/10.3390/atmos16050626, 2025.

Rothman, L., Rinsland, C., Goldman, A., Massie, S., Edwards, D., Flaud, J.-M., Perrin, A., Camy-Peyret, C., Dana, V., Mandin, J.-Y., Schroeder, J., Mccann, A., Gamache, R., Wattson, R., Yoshino, K., Chance, K., Jucks, K., Brown, L., Nemtchinov, V., and Varanasi, P.: THE HITRAN MOLECULAR SPECTROSCOPIC DATABASE AND HAWKS (HITRAN ATMOSPHERIC WORKSTATION): 1996 EDITION, Journal of Quantitative Spectroscopy and Radiative Transfer, 60, 665–710, https://doi.org/https://doi.org/10.1016/S0022-4073(98)00078-8, 1998.

SAFIRE ATR-42 aircraft: https://www.atr-aircraft.com/stories/safires-atr-42-a-flying-laboratory-revolutionising-atmospheric-research/, last access: 20 May 2026.

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., Dlugokencky, E. J., Etiope, G., Bastviken, D., Houweling, S., Janssens-Maenhout, G., Tubiello, F. N., Castaldi, S., Jackson, R. B., Alexe, M., Arora, V. K., Beerling, D. J., Bergamaschi, P., Blake, D. R., Brailsford, G., Brovkin, V., Bruhwiler, L., Crevoisier, C., Crill, P., Covey, K., Curry, C., Frankenberg, C., Gedney, N., Höglund-Isaksson, L., Ishizawa, M., Ito, A., Joos, F., Kim, H.-S., Kleinen, T., Krummel, P., Lamarque, J.-F., Langenfelds, R., Locatelli, R., Machida, T., Maksyutov, S., McDonald, K. C., Marshall, J., Melton, J. R., Morino, I., Naik, V., O'Doherty, S., Parmentier, F.-J. W., Patra, P. K., Peng, C., Peng, S., Peters, G. P., Pison, I., Prigent, C., Prinn, R., Ramonet, M., Riley, W. J., Saito, M., Santini, M., Schroeder, R., Simpson, I. J., Spahni, R., Steele, P., Takizawa, A., Thornton, B. F., Tian, H., Tohjima, Y., Viovy, N., Voulgarakis, A., van Weele, M., van der Werf, G. R., Weiss, R., Wiedinmyer, C., Wilton, D. J., Wiltshire, A., Worthy, D., Wunch, D., Xu, X., Yoshida, Y., Zhang, B., Zhang, Z., and Zhu, Q.: The global methane budget 2000–2012, Earth System Science Data, 8, 697–751, https://doi.org/10.5194/essd-8-697-2016, 2016.

Saunois, M., Martinez, A., Poulter, B., Zhang, Z., Raymond, P. A., Regnier, P., Canadell, J. G., Jackson, R. B., Patra, P. K., Bousquet, P., Ciais, P., Dlugokencky, E. J., Lan, X., Allen, G. H., Bastviken, D., Beerling, D. J., Belikov, D. A., Blake, D. R., Castaldi, S., Crippa, M., Deemer, B. R., Dennison, F., Etiope, G., Gedney, N., Höglund-Isaksson, L., Holgerson, M. A., Hopcroft, P. O., Hugelius, G., Ito, A.,

https://doi.org/10.5194/egusphere-2026-4050
Preprint. Discussion started: 11 August 2026

Jain, A. K., Janardanan, R., Johnson, M. S., Kleinen, T., Krummel, P. B., Lauerwald, R., Li, T., Liu, X., McDonald, K. C., Melton, J. R., Mühle, J., Müller, J., Murguia-Flores, F., Niwa, Y., Noce, S., Pan, S., Parker, R. J., Peng, C., Ramonet, M., Riley, W. J., Rocher-Ros, G., Rosentreter, J. A., Sasakawa, M., Segers, A., Smith, S. J., Stanley, E. H., Thanwerdas, J., Tian, H., Tsuruta, A., Tubiello, F. N., Weber, T. S., van der Werf, G. R., Worthy, D. E. J., Xi, Y., Yoshida, Y., Zhang, W., Zheng, B., Zhu, Q., Zhu, Q., and Zhuang, Q.: Global Methane Budget 2000–2020, Earth System Science Data, 17, 1873–1958, https://doi.org/10.5194/essd-17-1873-2025, 2025.

Schmidt, J., Wendisch, M., Curtius, J., Scheinert, M., and Sinnhuber, B.-M.: Über den Wolken, forschung, 43, 4–9, https://doi.org/https://doi.org/10.1002/fors.201870202, 2018.

Shaw, J. and Churnside, J.: Fractal laser glints from the ocean surface, Journal of the Optical Society of America A, 14, 1144–1150, https://doi.org/10.1364/JOSAA.14.001144, 1997a.

Shaw, J. A. and Churnside, J. H.: Scanning-laser glint measurements of sea-surface slope statistics, Appl. Opt., 36, 4202–4213, https://doi.org/10.1364/AO.36.004202, 1997b.

Shindell, D., Sadavarte, P., Aben, I., Bredariol, T. d. O., Dreyfus, G., Höglund-Isaksson, L., Poulter, B., Saunois, M., Schmidt, G. A., Szopa, S., Rentz, K., Parsons, L., Qu, Z., Faluvegi, G., and Maasakkers, J. D.: The methane imperative, Frontiers in Science, Volume 2 - 2024, https://doi.org/10.3389/fsci.2024.1349770, 2024.

Tellier, Y., Pierangelo, C., Wirth, M., Gibert, F., and Marnas, F.: Averaging bias correction for the future space-borne methane IPDA lidar mission MERLIN, Atmospheric Measurement Techniques, 11, 5865–5884, https://doi.org/10.5194/amt-11-5865-2018, 2018.

Weber, A., Kölling, T., Pörtge, V., Baumgartner, A., Rammeloo, C., Zinner, T., and Mayer, B.: Polarization upgrade of specMACS: calibration and characterization of the 2D RGB polarization-resolving cameras, Atmospheric Measurement Techniques, 17, 1419–1439, https://doi.org/10.5194/amt-17-1419-2024, 2024.