Atmospheric
Measurement
Techniques

Performance and sensitivity of column-wise and pixel-wise methane retrievals for imaging spectrometers

Alana K. Ayasse[1], **Daniel Cusworth**[1], **Kelly O'Neill**[1], **Justin Fisk**[1], **Andrew K. Thorpe**[2], **and Riley Duren**[1,2,3]

[1]Carbon Mapper Inc., Pasadena, CA 91101, USA
[2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
[3]Arizona Institutes for Resilience, University of Arizona, Tucson, AZ 85721, USA

Correspondence: Alana K. Ayasse (alana@carbonmapper.org)

Abstract. Strong methane point source emissions generate large atmospheric concentrations that can be detected and quantified with infrared remote sensing and retrieval algorithms. Two standard and widely used retrieval algorithms for one class of observing platform, imaging spectrometers, include pixel-wise and column-wise approaches. In this study, we assess the performance of both approaches using the airborne imaging spectrometer (Global Airborne Observatory) observations of two extensive controlled-release experiments. We find that the column-wise retrieval algorithm is sensitive to the flight line length and can have a systematic low bias with short flight lines, which is not present in the pixel-wise retrieval algorithm. However, the pixel-wise retrieval is very computationally expensive, and the column-wise retrieval algorithms can produce good results when the flight line length is sufficiently long. Lastly, this study examines the methane plume detection performance of the Global Airborne Observatory with a column-wise retrieval algorithm and finds minimum detection limits of between 9 of $10\,\mathrm{kg\,h^{-1}}$ and 90 % probability of detection between 10 and $45\,\mathrm{kg\,h^{-1}}$. These results present a framework of rules for guiding proper concentration retrieval selection given conditions at the time of observation in order to ensure robust detection and quantification.

1 Introduction

Multiple studies have shown that, in many regions across emission sectors, a significant component of the anthropogenic methane (CH_4) budget come from a relatively small population of high-emission discrete point sources (e.g., Lyon et al., 2015; Frankenberg et al., 2016; Irakulis-Loitxate et al., 2022; Lauvaux et al., 2022; Sherwin et al., 2023a; Duren et al., 2019; Cusworth et al., 2022). This result has significant policy implications, as identifying and mitigating a large proportion of CH_4 emissions quickly is required to limit adverse climate warming effects in the next few decades (Ocko et al., 2021). Several atmospheric remote-sensing platforms are particularly sensitive to these emission types. In particular, airborne imaging spectrometers with shortwave infrared (SWIR) sensitivity have emerged as useful tools for point source quantification due to their high spatial resolution, low detection limit, and ability to map large areas for point sources. The accuracy of the emission quantification of point sources depends on a combination of instrument performance, the CH_4 concentration retrieval algorithm, plume identification and delineation, and environmental variables including surface illumination and atmospheric transport (Gorroño et al., 2023; Ayasse et al., 2018). Methane retrieval algorithms for remote-sensing platforms that rely on solar backscattered radiance, like imaging spectrometers, vary in implementation and complexity (Jacob et al., 2022). As the ecosystem of airborne and satellite imaging spectrometers grows, understanding how retrieval assumptions propagate to emission estimates is required to ensure that remote-sensing-derived emission estimates are robust and accurate.

Controlled-release experiments provide a means of independently evaluating detection limits and uncertainty in emission estimates. Initial unblinded controlled-release experiments with the next-generation Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG) were performed to as-

sess detection limits and quantification accuracy at relatively low emission rates (maximum release was $141 \, \mathrm{kg} \, \mathrm{h}^{-1}$) (Thorpe et al., 2016). In 2021 and 2022, Stanford University conducted more comprehensive, blinded controlled-release experiments to validate multiple ground-based, airborne, and satellite CH_4 sensing technologies. Carbon Mapper, a non-profit organization that provides facility-scale CH_4 emission data via remote sensing, participated in both experiments. Carbon Mapper contracted the Global Airborne Observatory (GAO) imaging spectrometer, which has the same design as AVIRIS-NG, to collect the raw radiance data that Carbon Mapper then processed to emission estimates. The Carbon Mapper flights resulted in over 250 observations of metered emission rates between 5 and $1500 \, \mathrm{kg} \, \mathrm{h}^{-1}$. These data provide an excellent opportunity to test, compare, and validate CH_4 emissions retrieval algorithms.

In this paper, we use now unblinded controlled-release data to provide a quantitative sensitivity assessment of two CH_4 retrieval algorithms: (1) a column-wise matched filter algorithm and (2) a pixel-wise iterative maximum a posteriori – differential optical absorption spectroscopy (IMAP-DOAS) algorithm. We also use these data to assess a minimum detection limit for the Carbon Mapper airborne platform. During the 2021 and 2022 controlled-release experiments, we employed two observing strategies with GAO: (1) rapid repeat surveys focused on the controlled-release site, which resulted in many observations but smaller image sizes (2021 experiment), and (2) broad imaging of the controlled-release site and surrounding areas, which resulted in greater characterization of the background but fewer observations (2022 experiment). The latter strategy was more representative of our standard operations for mapping large regions. We show that the performance of the retrieval algorithm is strongly sensitive to the observing strategy, specifically for column-wise retrievals that depend on background characterization using many pixels across the scene. These results establish a general rule framework for selecting retrievals and quantifying systematic biases based on observing conditions at the time of acquisition. This is an especially critical finding given the computational efficiency, and therefore operational use, of the column-wise approach compared with the pixel-wise approach. These results corroborate a $10 \, \mathrm{kg} \, \mathrm{h}^{-1}$ detection limit for this class of airborne imaging spectrometer but also highlight the complexities and contributing factors that alter detection limits on a scene-by-scene bases. The rules and frameworks established here can be applied and adapted to other observing platforms that may apply similar CH_4 retrieval approaches.

2 Methods

2.1 Controlled-release experiments

A series of controlled-release experiments were performed by a Stanford University team in summer 2021 and fall 2022 (Rutherford et al., 2023; Sherwin et al., 2023b; El Abbadi et al., 2023). These experiments evaluated the detection limits and quantification accuracy of various CH_4 measurement technologies, including ground-based, airborne, and satellite platforms. In 2021, Carbon Mapper participated in controlled-release tests conducted on 30 and 31 July and on 3 August 2021 near Midland, Texas. The metered release rates were between 10 and $1500 \, \mathrm{kg} \, \mathrm{h}^{-1}$. In addition, a sonic anemometer was located at the site to provide wind speed and direction data at 10 m above ground level. In 2022, Carbon Mapper participated in controlled-release tests on 10–12, 28, 29, and 31 October 2022 near Casa Grande, Arizona. The metered release rates were between 5 and $1450 \, \mathrm{kg} \, \mathrm{h}^{-1}$, and 10 m sonic anemometer data were also provided. For more details on the controlled-release experiments, see Rutherford et al. (2023) (2021 experiments) and El Abbadi et al. (2023) (2022 experiments).

For both the 2021 and 2022 controlled releases, we used the GAO platform. The Visible to ShortWave Infrared (VSWIR) imaging spectrometer onboard GAO measures ground-reflected solar radiation at wavelengths from 380 to 2510 nm with 5 nm spectral sampling. The instrument spatial resolution is correlated to the flight altitude. For these studies, the instrument was flown at 10 000 ft ($\sim 3 \, \mathrm{km}$), resulting in an approximate pixel size of 3 m. GAO has a $34°$ field of view, which results in an approximate swath width (or cross-track extent) of approximately 2 km when flown at 3 km altitude. The flight line length refers to the along-track direction of data acquisition and can vary in length depending on observing preferences.

In the first controlled-release experiment, we maximized the number of observations during the short campaign window by flying short ($\sim 3 \, \mathrm{km}$) flight lines around the controlled-release site. This resulted in 229 observations of the controlled-release site (including null releases). However, this flight line length was well below our standard observing practice and had consequences with respect to the column-wise retrieval. Thus, for the second controlled-release experiment, the flight line length was extended to 20 km, which was more representative of normal survey operations, resulting in 121 observations of the controlled-release site. In both experiments, we applied our standard quality control protocols to eliminate scenes with clouds or cloud shadows, unstable plume morphology, or unstable wind conditions. After quality checks, there were 163 plumes for the 2021 controlled-release experiment and 87 plumes for the 2022 experiment.

2.2 Column-wise retrieval

The column matched filter (CMF) is a column-wise statistical algorithm to estimate CH_4 concentration enhancements. CMF algorithms are used operationally for Carbon Mapper airborne campaigns given their computational efficiency and ability to reduce systematic instrument error that may occur across nonuniform across-track elements in an imaging spectrometer. The algorithm takes the following form (Thompson et al., 2015):

$$\hat{\alpha}(x) = (x - \mu)^T \mathbf{\Sigma}^{-1} t / \left(t^T \mathbf{\Sigma}^{-1} t \right), \tag{1}$$

where $\hat{\alpha}$ is the path-length concentration CH_4 enhancement (units ppm m), x is a radiance spectrum, μ is the mean radiance spectrum in an along-track column, $\mathbf{\Sigma}$ is a covariance matrix, and t is a unit absorption spectrum. The vectors x, μ, and t include 71 elements, which represent all bands in the [2104, 2459] range, where CH_4 has known absorption properties (Roberts et al., 2010). In essence, the CMF uses statistics from all pixels in a flight column to assess whether a single pixel's spectrum has an CH_4 enhancement that is above the background concentration (i.e., has deeper absorption features). Sufficient column-wise statistics (i.e., pixels) are needed to define a robust covariance matrix. Initially, we supposed that the number of pixels in a column (i.e., flight line length) should be at least 7 times larger the number of active bands used in the retrieval and that no column should have more than 5 % of its pixels enhanced by CH_4. As we use 71 active bands in the retrieval, this suggests a $\sim$497-pixel flight line length minimum (1.5 km for a flight altitude of 3 km). Although this reasoning was used in general to define minimum flight line lengths, in practice, Carbon Mapper performs wide-area surveys of many facilities at regional and basin scales, resulting in much longer flight line lengths.

2.3 Pixel-wise retrieval

The IMAP-DOAS algorithm is a pixel-wise CH_4 retrieval. IMAP-DOAS estimates dry-air column-average CH_4 concentrations (XCH4; units ppb) on a per-pixel basis by simulation of top of the atmosphere (TOA) radiance and inversion (or retrieval) for the best atmospheric parameters that reduce mismatch between an observed spectrum and a simulated spectrum, assuming some prior constraints. For the simulated spectrum, IMAP-DOAS uses a radiative transfer model that relies on a multilayered Beer–Lambert law equation to simulate high-frequency atmospheric features and a multidimensional polynomial to represent low-frequency reflectance and scattering features (Cusworth et al., 2019; Frankenberg et al., 2005; Thorpe et al., 2017). The simulated spectrum must account for H_2O and N_2O, the other absorbing gases in the 2210–2400 nm spectra window, and low-frequency surface features modeled as a polynomial of order $k = [0, k]$. There-

fore, the state vector (x) is composed of the following:

$$x = \left(s_{CH_4}, s_{H_2O}, s_{N_2O}, a_0, \ldots, a_k \right), \tag{2}$$

where s is a scaling factor applied to the column mixing ratio for each gas from the US Standard Atmosphere. To retrieve the state vector from the radiance, we apply a forward model:

$$F^h(x) =$$
$$I_0(\lambda) \exp\left(-A \sum_{n=1}^{3} s_n \sum_{I=1}^{72} \tau_{n,l} \right) \sum_{k=0}^{K} \alpha_k P_k(\lambda), \tag{3}$$

where F^h is the high-resolution TOA radiance at wavelength λ; I_0 is the solar spectrum; A is the geometric air mass factor; $\tau_{n,l}$ is the optical depth for either CH_4, H_2O, or N_2O (n) and the vertical level (l); s_n is the scaling factor for that optical depth; α is a polynomial coefficient; and P is the kth polynomial. The optical depth $\tau_{n,l}$ is calculated at each wavelength by multiplying the HIgh-resolution TRANsmission (HITRAN) absorption cross section by the volume mixing ratio (VMR) and the vertical column density (VCD) of dry air in a 72-layer atmosphere from the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) meteorological reanalysis.

In order to model the TOA radiance, we take $F^h(X)$ over the 2210–2410 nm spectral range at a 0.02 nm resolution and convolve the spectrum using the band centers and full width at half maximum (FWHM) from the instrument (for AVIRIS-NG, this is a 5 nm spacing and a 6 nm FWHM). The observed TOA radiance (y) is represented as follows:

$$y = F(x) + \varepsilon, \tag{4}$$

where ε is the observational error. The forward model is nonlinear, so the solution must be obtained iteratively. At each iteration (i), a Jacobian matrix is calculated of the state vector.

$$\mathbf{K}_i = \frac{\partial F}{\partial x} \big|_{x=x_i} \tag{5}$$

We use Gauss–Newton iteration to solve for the optimal state vector:

$$x_{i+1} = x_A + \left(\mathbf{K}_i^T \mathbf{S}_O^{-1} \mathbf{K}_i + \mathbf{S}_A^{-1} \right)^{-1} \mathbf{K}_i^T \mathbf{S}_A^{-1}$$
$$\left[y - F(x_i) + \mathbf{K}_i (x_i - x_A) \right], \tag{6}$$

where $\mathbf{S}_O$ is the error covariance matrix defined by the signal-to-noise ratio (SNR), x_A is the prior estimate of the state vector, and $\mathbf{S}_A$ is the prior error covariance matrix.

IMAP-DOAS has been used in multiple previous studies for a smaller population of emission sources (Cusworth et al., 2020; Thorpe et al., 2017; Cusworth et al., 2021) but is currently not run operationally for large-area surveys due to computational constraints. The benefit of IMAP-DOAS is that (1) each retrieved pixel is independent and (2) retrieval

uncertainties can be explicitly characterized by the Bayesian formulation of the retrieval. This contrasts with the CMF approach, where the retrieved concentration for a single pixel depends on the quality and density of pixels in an acquisition (e.g., a single along-track column). However, the current processing time of IMAP-DOAS makes its operational use limited. At best, it takes 1 s per pixel to run; therefore, 5700 pixels (typically corresponding to a 300 m × 300 m area) can take 2–3 h to run. In contrast, it takes about 7 min to run an entire 3.3 million pixel scene with the CMF. Future processing improvements may significantly reduce the computation cost, but it is unlikely that IMAP-DOAS will ever be as computationally efficient as the CMF.

2.4 Emission rates

For both CMF and IMAP-DOAS, we estimate emission rates via the integrated methane enhancement (IME) method (Frankenberg et al., 2016; Varon et al., 2018):

$$Q = \frac{\text{IME}}{L} \times U_{\text{eff}}, \tag{7}$$

where Q is the emissions rate, IME is the integrated mass enhancement in kilograms, and L is the length in meters (in this case, the length is calculated using the square root of the area of the plume). U_{eff} is the effective wind speed and is calculated from the 10 m anemometer winds using the following equation:

$$U_{\text{eff}} = 1.1 \times \log(U) + 0.6, \tag{8}$$

where U is the 10 m anemometer wind speed. In the absence of a 10 m anemometer wind observation at the site of the plume, High-Resolution Rapid Refresh (HRRR) reanalysis products are used to estimate the 10 m wind speed at the time and location of the observed CH_4 plume.

The IME was calculated slightly differently for IMAP-DOAS and the CMF. For the CMF (which retrieves an enhancement above background and, therefore, does not have background CH_4 incorporated in the pixel values), we calculated thresholds using the 80th–98th percentile of the 1 km area around the plume origin. The thresholds were then used to filter out low values, and we subsequently calculated connected components starting at the plume origin and used a dilation to fill in gaps. The resulting plume mask was used to calculate the IME and L for both methods. In order to account for uncertainty due to the thresholds, this process was repeated for each percentile in the above range, and the IME and L values were averaged together to get the final IME and L. For IMAP-DOAS (which retrieves a total column concentration), we used the same methods stated above to determine the plume. However, IMAP-DOAS retrieves a total column concentration, so the background CH_4 needs to be subtracted off. The background concentration was determined by taking the pixels not included in the plume mask and taking a percentile (95th percentile for the 2021 plume
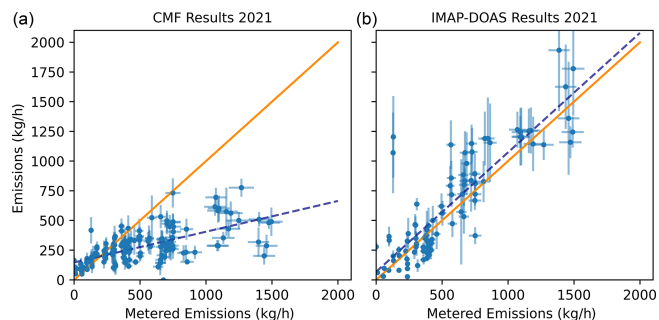

Figure 1. CMF **(a)** and IMAP-DOAS **(b)** comparison to metered emission rates for the 2021 controlled-release experiment with shorter-than-normal flight lines. An ordinary least squares (OLS) fit to CMF results in $y = 0.26x + 146$ with $R^2 = 0.42$. An OLS fit to IMAP-DOAS results in $y = 0.98x + 83$ with $R^2 = 0.72$. For both panels, the solid line represents the 1 : 1 line, the dashed line is the OLS fit, and the error bars represent 1σ uncertainties.

and 99th percentile for the 2022 plumes). For the CMF emission results, the uncertainties were derived from the standard deviation of the wind speeds 90 s prior to the observations and from the standard deviation of the IME and L due to different thresholds. For the IMAP-DOAS results, the uncertainties are derived directly from the retrieval and from the standard deviation of the winds as stated above.

3 Results and discussion

3.1 The 2021 controlled release

Estimated airborne emission rates compared to metered emissions are shown in Fig. 1 for both CMF and IMAP-DOAS. An ordinary least squares (OLS) fit between the CMF results and the metered emissions results in $y = 0.26x + 146$ and $R^2 = 0.42$. An OLS fit between IMAP-DOAS and metered emissions results in $y = 0.98x + 83$ and $R^2 = 0.72$. The CMF approach in this case significantly underestimated the metered emission rates. In contrast, emission rates derived from IMAP-DOAS showed good correlation and little bias across the population of releases. These values differ slightly from the values published in Rutherford et al. (2023) due to different emission quantification methods and quality filtering, but the trends remain the same.

The bias seen in the 2021 CMF result prompted us to revisit the line length assumption described above (i.e., the line length pixel minimum must be 7 times the length of active bands). A review of our standard line lengths for field campaigns, specifically the larger regional surveys of the Permian basin performed between 2019 and 2021 (Cusworth et al., 2019, 2022), revealed that the line lengths flown during the 2021 controlled-release experiment were an order of magnitude shorter than normal operations (Fig. 2). In the CMF formulation (Eq. 1), the magnitude of a CH_4 enhancement is di-
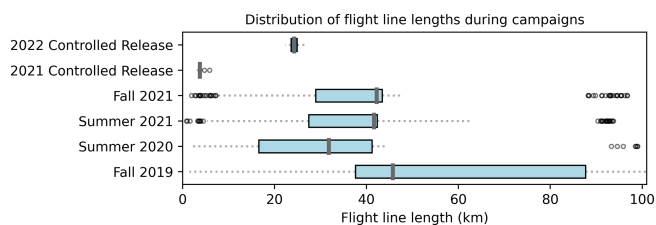
Figure 2. Flight line lengths during the 2021 and 2022 controlled-release experiments compared to Carbon Mapper Permian field campaigns (Cusworth et al., 2021, 2022). The data are displayed as a box plot, with the blue box representing the interquartile range, the gray bar representing the median, the dashed lines representing the respective minimum and maximum, and the black circles representing outliers.

Figure 3. Effect of cropping scene length on CMF results. Scenes were taken from 2019–2021 Permian campaigns that were flown under normal operations (i.e., not a controlled release experiment) and then cropped to 1200 pixels in length; following this, the CMF was rerun. The resulting emissions are compared to the emissions from the standard full-line CMF processing. The OLS fit results in $y = 0.35x + 40$ with $R^2 = 0.77$.

rectly related to the mean and covariance of pixels contained in a column. With a smaller flight line, the column covariance is calculated with a smaller number of pixels; this means that pixels with a CH_4 enhancement have a larger influence over the covariance, thereby making any deviations from the background (i.e., CH_4 enhancements) possibly smaller. The systematic low bias seen in Fig. 1a from the CMF result could therefore be indicative of flight lines that were systematically too short. In contrast, as IMAP-DOAS is a pixel-based algorithm and is therefore indifferent to flight lines lengths for quantification, the much closer agreement to metered emission rates in Fig. 1b is additional evidence that short flight lines drove much of the bias in the CMF result.

To provide further evidence that systematically short flight lines bias CMF-derived concentrations (and therefore emissions), we selected a subset of 50 flight lines that were flown during standard campaign operations in the Permian between 2019 and 2021. We isolated a single unique plume in each line, cropped the scene around that plume such that it was 1200 pixels (or 6 km) in the along-track direction, and then ran the CMF algorithm. As the 2019–2021 Permian campaigns were flown at higher altitudes (5–9 km), we cropped these scenes to match the average number of pixels per column during the 2021 controlled-release experiment. Therefore, these cropped scenes are representative of the statistical sampling conditions also present in the controlled-release CMF results. Figure 3 shows the results of cropping 2019–2021 Permian scenes to the same pixel dimension as the 2021 controlled-release experiment. What is immediately obvious is that estimated emissions from these cropped scenes are much lower than the standard CMF emission estimates that use all pixels in the along-track direction. An OLS fit between the cropped and standard CMF emissions results in $y = 0.14x + 81$ with $R^2 = 0.47$, showing a severe reduction in estimated emissions. This lends additional evidence that the poor results in Fig. 1 were driven primarily by short flight lines.

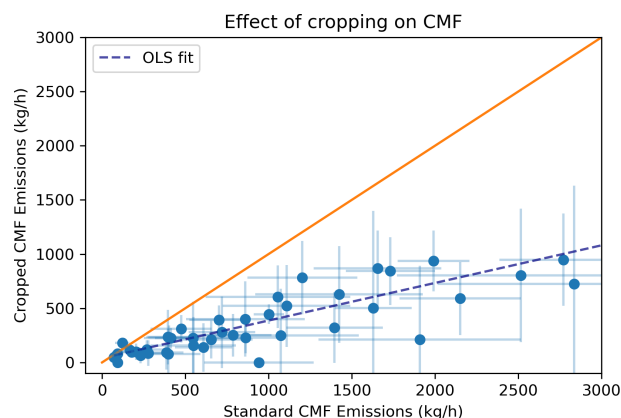The analyses described by Figs. 1, 2, and 3 lend confidence that the 2021 bias in CMF results was driven by

short flight lines. However, this unfortunately renders the 2021 controlled-release experiment incapable of assessing how a properly constrained (i.e., 20–50 km flight line length) CMF algorithm performs quantitatively against a standard metered emission rate. Nevertheless, given the good performance of IMAP-DOAS on the 2021 controlled-release data, we cross-compared emission rates from the 2019–2021 Permian campaigns derived from both CMF and IMAP-DOAS algorithms. This subset includes more than 60 plumes that relate to 20 distinct facilities that were imaged on at least 3 separate days during airborne campaigns by GAO during the Permian 2019–2021 campaigns (Cusworth et al., 2022). These plumes represent a dynamic range of emission rates reported by the CMF algorithm (90–3900 kg h^{-1}) and portray a diversity of infrastructure types in this region. The results of the IMAP-DOAS–CMF comparison are shown in Fig. 4. Figure 4a shows instantaneous plume-to-plume emission comparison for the different retrieval approaches. The data comparison shows general agreement between the two retrieval approaches. An OLS fit results in $y = 0.72x + 307$ with $R^2 = 0.67$.

In practice, when summarizing the results from campaigns and constructing emission budgets for regions or facilities, we take persistence-averaged emission rates derived from multiple overpasses of a facility (e.g., the average emission rate over multiple observations; Cusworth et al., 2021). Figure 4b shows the comparison of IMAP-DOAS and CMF after averaging and applying persistence adjustment to the emissions. Here, the comparison between retrieval approaches shows very close correspondence, with an OLS fit of $y = 0.89x + 120$ and $R^2 = 0.82$. As expected, the variability in emissions on a plume-by-plume basis gets averaged out in Fig. 4. As the only difference between CMF- and IMAP-
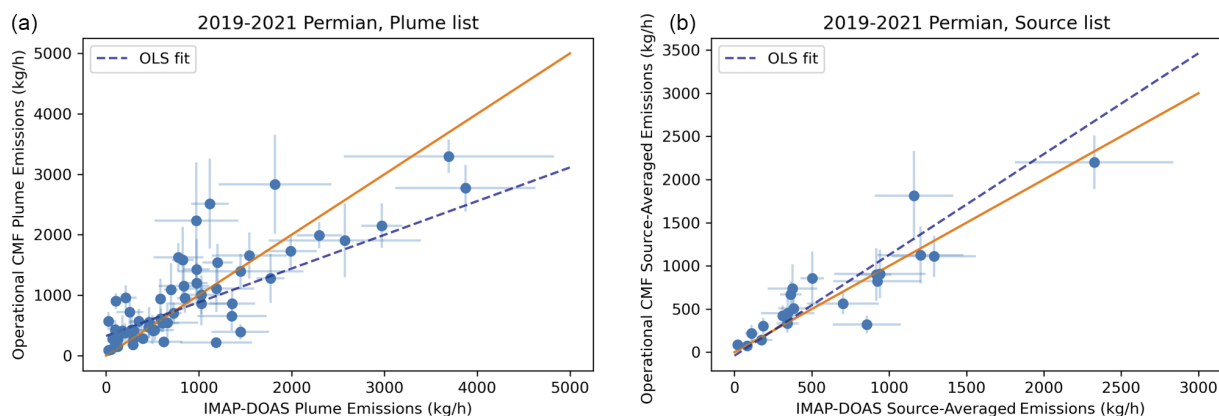
Figure 4. Panel **(a)** is a comparison of instantaneous emission rates for a subset of 2019–2021 plumes in the Permian Basin between the operational CMF and IMAP-DOAS. Panel **(b)** is a comparison of CMF and IMAP-DOAS for source-averaged emission rates. Error bars represent 1σ uncertainties in emissions. In panel **(a)**, the OLS regression fit (plume list with instantaneous emissions) is as follows: $y = 0.72x + 307$ with $R^2 = 0.67$.

DOAS-derived emission rates is the concentration retrievals, the improved correlation between single-realization plumes in Fig. 4a and multiple-realization sources in Fig. 4b shows that much of the retrieval uncertainty from these two approaches is unbiased because they begin to converge in agreement with additional sampling.

3.2 Column matched filter sensitivity tests

The results in Figs. 1 and 3 provide evidence that a reduced flight line length hampers the CMF's quantitative performance; however, the results do not reveal what the appropriate line length is to ensure good quantification. In order to improve results for the 2022 controlled release and to ensure that we are able to accurately quantify plumes operationally, we performed an analysis to determine the minimum line length for good quantification. To do this we identified eight sources that had at least three flyovers from GAO and were located in long lines. This resulted in 24 individual scenes to analyze. These scenes ranged from 125 to 27 km in length. Using the full flight line lengths, the plumes within each scene have CMF-derived emission rates between 70.7 and 3980 kg h^{-1}. Scenes were acquired from various campaigns, including the Permian 2021, Denver-Julesburg summer 2021, North East 2021, California summer 2020, and Permian fall 2021 (Thorpe et al., 2023; Cusworth et al., 2022).

Each scene was cropped to 2000 pixels (6–10 km) centered on the identified plume. This crop was then iteratively increased in 1000-pixel (3–5 km) increments until the full scene length was reached. This produced between 6 and 24 cropped images per scene. Each scene crop was processed through the CMF and an IME algorithm. We used IME over emission rates because the IME isolated changes in the CH$_4$ retrieval better than a full emission estimate. Additional de-

tails on the methods used in this analysis can be found in Sect. S3 in the Supplement.

We found that, in general, the IME increases as the pixel count in each column increases; however, the rate of increase decreases with pixel count. While not asymptotic, the small increase in the IME after a certain threshold likely means that there is an optimal point (or distance) that is sufficient for quantifying an emission rate. Figure 5a shows the pixel count vs. the IME for three different images of one CH$_4$ plume. The IME increases rapidly at the low pixel counts (short lines) and then levels off as the line gets longer. We determined this optimal point by calculating the "knee" in the curve (or the point where the change in IME is minimal compared with the change in the scene length). We found that, for all the plumes assessed, the median knee was 7000 pixels (or about 21 km when the aircraft is flown as 3 km). We compared the IME from the 7000 pixels to the standard scene length IME and found generally good agreement (Fig. 5b). From this, we can also conclude that the minimum line length needed to produce a good quantification with the CMF is about 21 km. This minimum line length is well within the lengths flown from previous Permian surveys (Fig. 2).

3.3 The 2022 controlled release

The results in Fig. 5 prompted us to require a minimum flight line distance of 20 km during the 2022 controlled-release experiment. The estimated airborne emission rates compared to metered emissions for the 2022 controlled release are shown in Fig. 6 for both CMF and IMAP-DOAS. The OLS fit between the CMF results and the metered emissions results in $y = 0.90x + 42$ and $R^2 = 0.88$. The OLS fit between IMAP-DOAS and metered emissions results in $y = 1.14x - 19$ and $R^2 = 0.81$. These values differ slightly from the values published in El Abbadi et al. (2023) due to different quality filters, but the trends remain the same. The results of the CMF
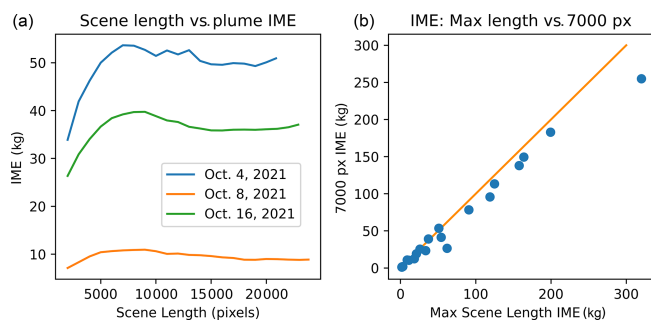
Figure 5. Panel **(a)** shows the scene length (pixels) vs. the IME from the matched filter for three plumes from a campaign in October 2021. Panel **(b)** shows the comparison of calculated emission rate plumes with a scene length of 7000 pixels (or 21 km when the aircraft is flown at 3 km) compared to the calculated emission rate for the full scene length (length variable depending on scene). The agreement is good, indicating that a ~ 20 km line is a sufficient length for good quantification.
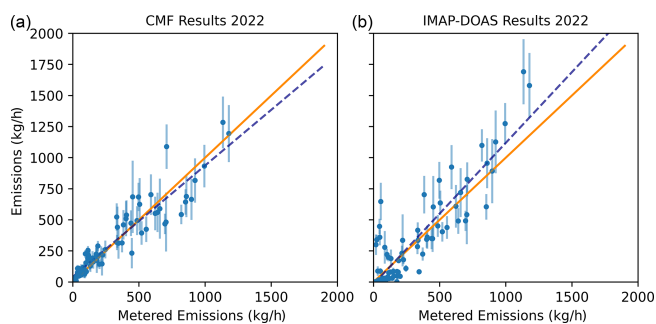

Figure 6. CMF and IMAP-DOAS comparison to metered emission rates for the 2022 controlled-release experiment. An OLS fit to CMF results in $y = 0.90x + 42$ with $R^2 = 0.88$. An OLS fit to IMAP-DOAS results in $y = 1.14x - 19$ with $R^2 = 0.81$. For both panels, the solid line represents the $1 : 1$ line and the dashed line is the OLS fit. The error bars represent 1σ uncertainties; there is error present on the x axis, but the error is too small to visualize.

are much improved from the 2021 controlled release. These results further highlight the need for appropriately long flight lines. The improved results provide additional assurance that previous airborne campaigns, most of which have flight lines longer than 20 km, do not have a systematic underestimate.

The CMF and IMAP-DOAS produced comparable results for the 2022 experiment, similar to the results seen from Permian campaigns shown in Fig. 4. However, other features exist that may corrupt a CMF result, even for long flight lines that were not explicitly tested during the 2022 controlled-release experiment. For example, flares or specular reflections from solar panels often result in saturated or atypical backscattered radiance spectra. If these spectra are observed anywhere in a column and not removed, they can enter into a covariance calculation, causing the CMF results in that column to be biased. In addition, the presence of too many

dark pixels, like a waterbody, can also negatively affect the CMF. Although not affected by other pixels, physics-based retrievals like IMAP-DOAS are also prone to retrieval artifacts for complicated pixels. Therefore, controlled-release tests across a host of simple to complex observing environments will further refine algorithms to quantify concentrations and emissions.

3.4 Detection limits

Although biased by short flight lines, the CMF results from the 2021 controlled-release experiment performed well with respect to CH_4 plume detection. The CMF ability to detect emissions has proven insensitive to line length, but these detection limits are strongly influenced by other observing conditions. To determine detection limits, we used logistic regression to build detection models. For the first set of models (Fig. 7a), we only used the metered emission rates as the predictor variable and built separate models for the 2021 and 2022 experiments. This allows us to compare the detection limits under two different observing conditions. We also made a more general model (Fig. 7b); for this, we combined both experiments and used wind speed as well as the metered emission rates as predictor variables. Other factors like solar zenith angle (SZA) and albedo should also be used as predictor variables; however, for these experiments, there was not enough variability in albedo to use it as a predictor variable, and the SZA, in this case driven by time of day, was also correlated with the metered release rates and was, therefore, not used as a predictor variable.

In the 2022 controlled-release experiment, the smallest plume detected was $8.6\,\mathrm{kg\,h^{-1}}$ and the 90 % probability of detection (POD) was $10\,\mathrm{kg\,h^{-1}}$. However, these were ideal conditions. The albedo was 42 % in the area surrounding the controlled-release site, which is considered high, and the wind speeds were low (mean of $1.78\,\mathrm{m\,s^{-1}}$). Typical operating conditions are more challenging, which can lead to higher minimum detection limits. The 2021 controlled release had more challenging conditions: the wind speeds were higher (mean of $2.5\,\mathrm{m\,s^{-1}}$), although not extreme, and the surface albedo was 24.5 % and also more varied. In the 2021 experiment, the smallest plume detected was $9.8\,\mathrm{kg\,h^{-1}}$ and the 90 % POD was $45\,\mathrm{kg\,h^{-1}}$. It is important to note that the smallest release rate was also $9.8\,\mathrm{kg\,h^{-1}}$; thus, for 2021, there are no data for releases below this value, and our understanding of detection below $9.8\,\mathrm{kg\,h^{-1}}$ is consequently incomplete. When we combine both controlled-release experiments and add wind as a predictor variable, we can more clearly visualize how detection limits can change depending on the conditions. The 90 % POD ranges from 17 to $68\,\mathrm{kg\,h^{-1}}$ as a function of wind speed (Fig. 7). These results are consistent with previous controlled-release results and with other analyses of POD curves that showed an AVIRIS-NG 90 % POD of $16–33\,\mathrm{kg\,h^{-1}}$ (Thorpe et al., 2016; Conrad et al., 2023). The POD is also a function of albedo and SZA, but these
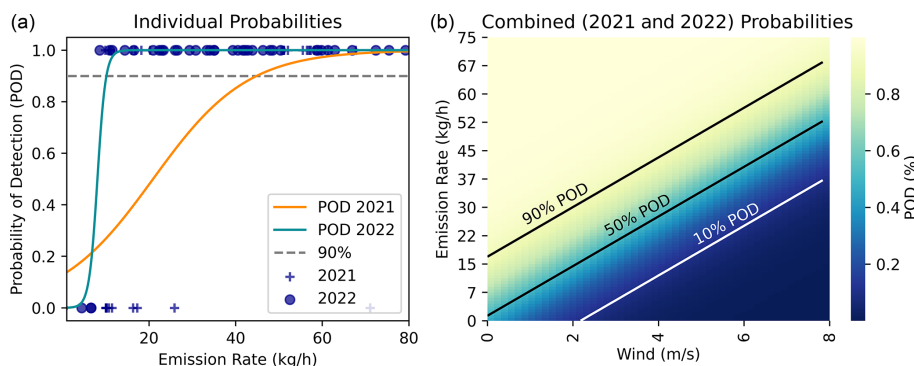
Figure 7. Probability of detection (POD) for the AVIRIS-NG/GAO instrument using the CMF methane retrieval algorithm. Panel **(a)** shows the individual POD curves for the 2021 and 2022 controlled release. The points at zero represent null detections and the points at one represent positive detections. Panel **(b)** shows the combined POD contours and includes wind.

controlled-release experiments were not designed to test and isolate those parameters. In practice, we anticipate the POD performance to vary across observing regions and seasons (Gorroño et al., 2023). Specifying and understanding observing conditions is critical to interpreting POD and minimum detection results.

4 Conclusion

The nonstandard short flight lines flown during the 2021 controlled-release experiment resulted in an unexpected low bias in CMF-retrieved CH_4 concentrations which propagated into low emission rates. Here, we tested that hypothesis by applying the pixel-wise IMAP-DOAS retrieval to the 2021 controlled-release results. We found a much improved result, with IMAP-DOAS-derived emission rates showing strong correlation and little bias across the experiment. We subsequently flew longer, more representative flight lines in the 2022 controlled-release experiment and eliminated bias in both the CMF and IMAP-DOAS retrievals. Both experiments were used to assess the minimum detection limit as well as the 90 % POD. These experiments highlighted the importance of observing conditions when evaluating the POD.

Here, we also highlight the various strength and weaknesses of the two main retrieval algorithms. The CMF is a fast and reliable detection algorithm but is sensitive to scene dynamics, specifically scene length, for quantification. Here, we show that the 2021 controlled release did not meet the scene length requirements for good quantification, but we also show that most Carbon Mapper flight lines meet the scene length requirement and produce good quantitative results. IMAP-DOAS, on the other hand, is too slow to run as a detection algorithm, but it can produce more reliable quantification estimates and does not rely on other aspects of a scene. However, continuing work is needed to robustly quantify background concentrations over a diverse set of observing conditions. Ideally, these two algorithms can be used in

tandem: the CMF for rapid detection and IMAP-DOAS after for a robust quantification.

As we move towards routine and global monitoring with remote sensing, these results provide confidence in our ability to accurately quantify CH_4 emissions. These results also highlight the importance of controlled-release testing in order to assess and understand retrieval algorithms. As a larger constellation of instruments are used to map CH_4 (Jacob et al., 2022), more controlled-release tests will be needed to fully validate emissions.

Review statement. This paper was edited by Christoph Kiemle and reviewed by two anonymous referees.

References

Ayasse, A. K., Thorpe, A. K., Roberts, D. A., Funk, C. C., Dennison, P. E., Frankenberg, C., Steffke, A., and Aubrey, A. D.: Evaluating the effects of surface properties on methane retrievals using a synthetic airborne visible/infrared imaging spectrometer next generation (AVIRIS-NG) image, Remote Sens. Environ., 215, 386–397, https://doi.org/10.1016/j.rse.2018.06.018, 2018.

Carbon Mapper: Map, Carbon Mapper, OpenStreetMap [data set], https://data.carbonmapper.org (last access: 20 April 2022), 2021.

Conrad, B. M., Tyner, D. R., and Johnson, M. R.: Robust probabilities of detection and quantification uncertainty for aerial methane detection: Examples for three airborne technologies, Remote Sens. Environ., 288, 113499, https://doi.org/10.1016/j.rse.2023.113499, 2023.

Cusworth, D. H., Jacob, D. J., Varon, D. J., Chan Miller, C., Liu, X., Chance, K., Thorpe, A. K., Duren, R. M., Miller, C. E., Thompson, D. R., Frankenberg, C., Guanter, L., and Randles, C. A.: Potential of next-generation imaging spectrometers to detect and quantify methane point sources from space, Atmos. Meas. Tech., 12, 5655–5668, https://doi.org/10.5194/amt-12-5655-2019, 2019.

Cusworth, D. H., Duren, R. M., Yadav, V., Thorpe, A. K., Verhulst, K., Sander, S., Hopkins, F., Rafiq, T., and Miller, C. E.: Synthesis of methane observations across scales: Strategies for deploying a multitiered observing network, Geophys. Res. Lett., 47, e2020GL087869, https://doi.org/10.1029/2020gl087869, 2020.

Cusworth, D. H., Duren, R. M., Thorpe, A. K., Pandey, S., Maasakkers, J. D., Aben, I., Jervis, D., Varon, D. J., Jacob, D. J., Randles, C. A., Gautam, R., Omara, M., Schade, G. W., Dennison, P. E., Frankenberg, C., Gordon, D., Lopinto, E., and Miller, C. E.: Multisatellite imaging of a gas well blowout enables quantification of total methane emissions, Geophys. Res. Lett., 48, e2020GL090864, https://doi.org/10.1029/2020gl090864, 2021.

Cusworth, D. H., Thorpe, A. K., Ayasse, A. K., Stepp, D., Heckler, J., Asner, G. P., Miller, C. E., Yadav, V., Chapman, J. W., Eastwood, M. L., Green, R. O., Hmiel, B., Lyon, D. R., and Duren, R. M.: Strong methane point sources contribute a disproportionate fraction of total emissions across multiple basins in the United States, P. Natl. Acad. Sci. USA, 119, e2202338119, https://doi.org/10.1073/pnas.2202338119, 2022.

Duren, R. M., Thorpe, A. K., Foster, K. T., Rafiq, T., Hopkins, F. M., Yadav, V., Bue, B. D., Thompson, D. R., Conley, S., Colombi, N. K., Frankenberg, C., McCubbin, I. B., Eastwood, M. L., Falk, M., Herner, J. D., Croes, B. E., Green, R. O., and Miller, C. E.: California's methane super-emitters, Nature, 575, 180–184, https://doi.org/10.1038/s41586-019-1720-3, 2019.

El Abbadi, S., Chen, Z., Burdeau, P., Rutherford, J., Chen, Y., Zhang, Z., Sherwin, E., and Brandt, A.: Comprehensive evaluation of aircraft-based methane sensing for greenhouse gas mitigation, EarthArXiv, https://doi.org/10.31223/x51d4c, 2023.

Frankenberg, C., Platt, U., and Wagner, T.: Iterative maximum a posteriori (IMAP)-DOAS for retrieval of strongly absorbing trace gases: Model studies for CH_4 and CO_2 retrieval from near infrared spectra of SCIAMACHY onboard ENVISAT, Atmos. Chem. Phys., 5, 9–22, https://doi.org/10.5194/acp-5-9-2005, 2005.

Frankenberg, C., Thorpe, A. K., Thompson, D. R., Hulley, G., Kort, E. A., Vance, N., Borchardt, J., Krings, T., Gerilowski, K., Sweeney, C., Conley, S., Bue, B. D., Aubrey, A. D., Hook, S., and Green, R. O.: Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region, P. Natl. Acad. Sci. USA, 113, 9734–9739, https://doi.org/10.1073/pnas.1605617113, 2016.

Gorroño, J., Varon, D. J., Irakulis-Loitxate, I., and Guanter, L.: Understanding the potential of Sentinel-2 for monitoring methane point emissions, Atmos. Meas. Tech., 16, 89–107, https://doi.org/10.5194/amt-16-89-2023, 2023.

Irakulis-Loitxate, I., Guanter, L., Maasakkers, J. D., Zavala-Araiza, D., and Aben, I.: Satellites Detect Abatable Super-Emissions in One of the World's Largest Methane Hotspot Regions, Environ. Sci. Technol., 56, 2143–2152, https://doi.org/10.1021/acs.est.1c04873, 2022.

Jacob, D. J., Varon, D. J., Cusworth, D. H., Dennison, P. E., Frankenberg, C., Gautam, R., Guanter, L., Kelley, J., McKeever, J., Ott, L. E., Poulter, B., Qu, Z., Thorpe, A. K., Worden, J. R., and Duren, R. M.: Quantifying methane emissions from the global scale down to point sources using satellite observations of atmospheric methane, Atmos. Chem. Phys., 22, 9617–9646, https://doi.org/10.5194/acp-22-9617-2022, 2022.

Lauvaux, T., Giron, C., Mazzolini, M., d'Aspremont, A., Duren, R., Cusworth, D., Shindell, D., and Ciais, P.: Global assessment of oil and gas methane ultra-emitters, Science, 375, 557–561, https://doi.org/10.1126/science.abj4351, 2022.

Lyon, D. R., Zavala-Araiza, D., Alvarez, R. A., Harriss, R., Palacios, V., Lan, X., Talbot, R., Lavoie, T., Shepson, P., Yacovitch, T. I., Herndon, S. C., Marchese, A. J., Zimmerle, D., Robinson, A. L., and Hamburg, S. P.: Constructing a Spatially Resolved Methane Emission Inventory for the Barnett Shale Region, 49, 8147–8157, https://doi.org/10.1021/es506359c, 2015.

Ocko, I. B., Sun, T., Shindell, D., Oppenheimer, M., Hristov, A. N., Pacala, S. W., Mauzerall, D. L., Xu, Y., and Hamburg, S. P.: Acting rapidly to deploy readily available methane mitigation measures by sector can immediately slow global warming, Environ. Res. Lett., 16, 054042, https://doi.org/10.1088/1748-9326/abf9c8, 2021.

Roberts, D. A., Bradley, E. S., Cheung, R., Leifer, I., Dennison, P. E., and Margolis, J. S.: Mapping methane emissions from a marine geological seep source using imaging spectrometry, Remote Sens. Environ., 114, 592–606, https://doi.org/10.1016/j.rse.2009.10.015, 2010.

Rutherford, J., Sherwin, E., Chen, Y., Aminfard, S., and Brandt, A.: Evaluating methane emission quantification performance and uncertainty of aerial technologies via high-volume single-blind controlled releases, EarthArXiv, https://doi.org/10.31223/x5kq0x, 2023.

Sherwin, E., Rutherford, J., Zhang, Z., Chen, Y., Wetherley, E., Yakovlev, P., Berman, E., Jones, B., Thorpe, A., Ayasse, A., Duren, R., Brandt, A., and Cusworth, D.: Quantifying oil and natural gas system emissions using one million aerial site measurements, Research Square, https://doi.org/10.21203/rs.3.rs-2406848/v1, 2023a.

Sherwin, E. D., Rutherford, J. S., Chen, Y., Aminfard, S., Kort, E. A., Jackson, R. B., and Brandt, A. R.: Single-blind validation of space-based point-source detection and quantification of onshore methane emissions, Sci. Rep., 13, 1–10, https://doi.org/10.1038/s41598-023-30761-2, 2023b.

Thompson, D. R., Leifer, I., Bovensmann, H., Eastwood, M., Fladeland, M., Frankenberg, C., Gerilowski, K., Green, R. O., Kratwurst, S., Krings, T., Luna, B., and Thorpe, A. K.: Real-time remote detection and measurement for airborne imaging spectroscopy: a case study with methane, Atmos. Meas. Tech., 8, 4383–4397, https://doi.org/10.5194/amt-8-4383-2015, 2015.

Thorpe, A. K., Frankenberg, C., Aubrey, A. D., Roberts, D. A., Nottrott, A. A., Rahn, T. A., Sauer, J. A., Dubey, M. K., Costigan, K. R., Arata, C., Steffke, A. M., Hills, S., Haselwimmer, C., Charlesworth, D., Funk, C. C., Green, R. O., Lundeen, S. R., Boardman, J. W., Eastwood, M. L., Sarture, C. M., Nolte, S. H., Mccubbin, I. B., Thompson, D. R., and McFadden, J. P.: Mapping methane concentrations from a controlled release experiment using the next generation airborne visible/infrared imaging spectrometer (AVIRIS-NG), Remote Sens. Environ., 179, 104–115, https://doi.org/10.1016/j.rse.2016.03.032, 2016.

Thorpe, A. K., Frankenberg, C., Thompson, D. R., Duren, R. M., Aubrey, A. D., Bue, B. D., Green, R. O., Gerilowski, K., Krings, T., Borchardt, J., Kort, E. A., Sweeney, C., Conley, S., Roberts, D. A., and Dennison, P. E.: Airborne DOAS retrievals of methane, carbon dioxide, and water vapor concentrations at high spatial resolution: application to AVIRIS-NG, Atmos. Meas. Tech., 10, 3833–3850, https://doi.org/10.5194/amt-10-3833-2017, 2017.

Thorpe, A. K., Kort, E. A., Cusworth, D. H., Ayasse, A. K., Bue, B. D., Yadav, V., Thompson, D. R., Frankenberg, C., Herner, J., Falk, M., Green, R. O., Miller, C. E., and Duren, R. M.: Methane emissions decline from reduced oil, natural gas, and refinery production during COVID-19, Environ. Res. Commun., 5, 021006, https://doi.org/10.1088/2515-7620/acb5e5, 2023.

Varon, D. J., Jacob, D. J., McKeever, J., Jervis, D., Durak, B. O. A., Xia, Y., and Huang, Y.: Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes, Atmos. Meas. Tech., 11, 5673–5686, https://doi.org/10.5194/amt-11-5673-2018, 2018.