Methane Point Source Quantification Using MethaneAIR: A New Airborne Imaging Spectrometer

Apisada Chulakadabba[1], Maryann Sargent[1], Thomas Lauvaux[2], Joshua S. Benmergui[1, 3, 4], Jonathan E. Franklin[1], Christopher Chan Miller[1, 5], Jonas S. Wilzewski[1, 5], Sébastien Roche[1, 5], Eamon Conway[5, 6], Amir H. Souri[5], Kang Sun[7, 8], Bingkun Luo[5], Jacob Hawthrone[5], Jenna Samra[5], Bruce C. Daube[1], Xiong Liu[5], Kelly Chance[5], Yang Li[9], Ritesh Gautam[3, 4], Mark Omara[3, 4], Jeff S. Rutherford[10], Evan D. Sherwin[10], Adam Brandt[10], and Steven C. Wofsy[1]

[1]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
[2]Molecular and Atmospheric Spectrometry Group (GSMA) – UMR 7331, University of Reims Champagne Ardenne, France
[3]Environmental Defense Fund, New York, NY
[4]MethaneSAT, LLC, Austin, TX
[5]Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA
[6]Kostas Research Institute for Homeland Security, Northeastern University, Burlington, MA, USA
[7]Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, NY, USA
[8]Research and Education in Energy, Environment and Water Institute, University at Buffalo, Buffalo, NY, USA
[9]Department of Environmental Science, Baylor University, Waco, TX
[10]Department of Energy Science & Engineering, Stanford University, Stanford, CA

Correspondence: Apisada Chulakadabba (achulakadabba@seas.harvard.edu)

Abstract. The MethaneSAT satellite instrument and its aircraft precursor, MethaneAIR, are imaging spectrometers designed to measure methane concentrations with wide spatial coverage, fine spatial resolution, and high precision compared to currently deployed remote sensing instruments. At 12960 m cruise altitude above ground (13850 above sea level), MethaneAIR datasets have a 4.5 km swath gridded to 10m x 10m pixels with 17 - 20 ppb standard deviation on a flat scene. ~~It~~ MethaneAIR was deployed in the summer of 2021 in the Permian Basin to test the accuracy of the retrieved methane concentrations and emission rates using the algorithms developed for MethaneSAT. We report here point source emissions obtained during a single-blind volume controlled release experiment, using two methods: (1) The modified Integrated Mass Enhancement (mIME) method estimates emission rates using the total mass enhancement of methane in an observed plume combined with winds obtained from Weather Research Forecast driven by High-Resolution Rapid Refresh meteorological data in Large Eddy Simulations mode (WRF-LES-HRRR). WRF-LES-HRRR simulates winds in stochastic eddy-scale (100 - 1000 m) variability, which is particularly important for low-wind conditions and informing the error budget. The mIME can estimate emission rates of plumes of any size that are detectable by MethaneAIR. (2) The Divergence Integral (DI) method applies Gauss's theorem to estimate the flux divergence fields through a series of closed surfaces enclosing the sources. The set of boxes grows from the upwind side of the plume through the core of each plume and downwind. No selection of inflow concentration, as used in the mIME, is required. The DI approach can efficiently determine fluxes from large sources and clusters of sources but cannot resolve small point emissions. These methods account for the effects of eddy-scale variation in different ways: the DI averages across many eddies, whereas the mIME re-samples many eddies from the LES simulation~~; they also use different~~

The DI directly uses HRRR winds, while mIME uses WRF-LES-HRRR wind products. Emissions estimates from both the mIME and DI methods agreed closely with the blinded-volume controlled releases experiments ($N = 21$). The York regression between the estimated emissions and the released emissions has a slope of 0.96 [0.84, 1.08], $R = 0.83$ and $N = 21$, with 30% mean percentage error for the whole data set, which indicates that MethaneAIR can quantify point sources emitting more than 200 kg/hr for the mIME and 500 kg/hr for the DI method. The two methods also agreed on methane emission estimates from various uncontrolled sources in the Permian Basin. The experiment thus demonstrates the powerful potential of ~~our instruments for remote sensing and quantification of methane emissions~~the MethaneAIR instrument and suggests that the quantification method should be transferable to MethaneSAT if it meets the design specifications.

1   Introduction

Methane (CH_4) is the second most important anthropogenic greenhouse gas, with more than 80 times the warming potential of carbon dioxide (CO_2) in the first 20 years after its release (Myhre et al., 2013; Etminan et al., 2016). Due to its shorter atmospheric lifetime and higher thermal infrared absorbing efficiency, reduction of methane emissions offers an attractive option for near-term mitigation of greenhouse gas emissions (Shindell et al., 2012). Since the oil and gas (O&G) industry accounts for more than 22% of anthropogenic methane emissions, reducing O&G methane emissions is an effective strategy for reducing greenhouse gas emissions with the added benefit of reduced loss of valuable natural gas (Saunois et al., 2020). Identifying and quantifying methane emissions over large regions with high accuracy is crucial for achieving this goal.

Motivated by these concerns, several remote sensing instruments designed to image high concentrations of methane very close to point sources ("point source imagers") have been introduced in the last decade, including hyperspectral airborne systems such as AVIRIS-NG of Carbon Mapper (Frankenberg et al., 2016) ~~,~~ (not explicitly designed for methane measurement), Ball Aerospace' Methane Monitor (Bartholomew et al., 2017), Kairos (Sherwin et al., 2021), Bridger Photonics (Johnson et al., 2021), and satellites including PRISMA (Guanter et al., 2021), Sentinel-2 (Varon et al., 2021), and WorldView-3 (Sánchez-García et al., 2021) (Sentinel-2 or WorldView-3 were also not designed for methane measurements), plus a satellite interferometer, GHGSat (Jervis et al., 2021). There ~~is also a global mapping satellite~~are also global mapping satellites, TROPOspheric Monitoring Instrument (TROPOMI) (Veefkind et al., 2012), with the capability to map the entire globe daily at a much lower spatial resolution (5km x 7km) and The MEthane Remote sensing Lidar missioN (MERLIN) (Ehret et al., 2017) at 200km by 200km resolution.

The Environmental Defense Fund (EDF) and its wholly-owned subsidiary, MethaneSAT LLC, have developed a satellite mission, MethaneSAT, to fill the gap between these remote sensing approaches. MethaneAIR is an aircraft-based precursor of MethaneSAT, using ~~nearly identical~~ very similar spectroscopy, for developing and validating MethaneSAT algorithms (Conway et al., 2023; Staebell et al., 2021). Both instruments were designed to achieve wide spatial coverage with fine spatial resolution and high precision compared to existing remote sensing instruments. MethaneSAT, scheduled for launch in Q1 of 2024, ~~will~~ is designed to have the capability to quantify methane emissions at regional scales (10 km - 100 km), including diffuse emissions as well as detailed resolution of point sources. MethaneAIR, currently deployed, can also provide high-

resolution large-spatial scale methane emission detection and quantification as a stand-alone instrument ~~(Chan Miller et al., 2022)~~ .

This paper presents the results from the first flights of MethaneAIR, focused on ~~testing and validating retrievals of methane concentrations and~~ validating emission estimates for O&G sources based on MethaneAIR retrievals. Prior remote sensing studies have introduced several methods to estimate tracer emissions from point sources using methane concentration data. Methods have included applications of Gauss's divergence theorem, including the divergence integral (DI) approach (Frankenberg et al., 2016), mass flux (Conley et al., 2016), or cross-sectional flux methods (Varon et al., 2018). These methods assume that the flux across an enclosing surface around the source equals the integral of flux divergence over the same enclosing surface. The accuracy of these methods is limited by the accuracy of the wind profiles used, the spatial resolution and precision of methane observations, knowledge of the vertical distribution of methane and boundary conditions, and the influence of nearby methane sources. A source pixel method can be used when resolution is coarse (up to at least $10 \times 10\ km^2$), but measured sensitivity is high (Jacob et al., 2016). This method is based on observations very close to the point emission, allowing the observer to neglect information from the plume downwind where turbulent motions may lead to complex morphology. More recent approaches such as integrated mass enhancement (IME) and machine learning (Varon et al., 2018; Jongaramrungruang et al., 2019) combine the measured excess mass of methane near a source, compared to inflow concentrations, with an effective wind velocity flushing the excess downwind, to obtain an emission estimate (discussed in detail below). These methods work best with a fine spatial resolution (≤ 1 km) and high measurement sensitivity (column precision $\leq 10\,\%$). However, cases with multiple point sources in close proximity to one another are still challenging for all the methods.

We carried out a single-blind volume-controlled release experiment to verify the methods for quantifying methane concentrations and point source emissions using MethaneAIR data. The ground team (the Stanford team) released methane at various metered emission rates ~~dein~~ from a site near Midland, TX, without sharing the information with the MethaneAIR team. The MethaneAIR team repeatedly flew over the site on two different days, and estimated emission rates of methane without prior information on actual emission rates or measured wind speeds, and submitted the results to the Stanford team. The Stanford team then revealed the volume-controlled release rates. The MethaneAIR team then made minor adjustments to their retrieval algorithms and quality control framework ("decision tree"). This procedure is similar to that used in prior remote sensing blinded-volume controlled release efforts including Sherwin et al. (2021), Frankenberg et al. (2016), Sherwin et al. (2022), and Johnson et al. (2021).

Our methods build upon previous IME and DI algorithms that we adapted to take advantage of MethaneAIR's high resolution and large spatial coverage. We used the Weather Research Forecast model version 3.9.1 in Large Eddy Simulation mode combined with geographical and mesoscale data (HRRR meteorological fields, see below) to simulate eddy-scale winds and vertical mixing. Our application of a high-resolution meteorological (LES) model provides important information for source estimation at the scales measured by MethaneAIR. We validated our results by (1) using the single blinded-volume controlled release experiment and (2) comparing our modified integrated mass enhancement (mIME) and DI methods when applied to uncontrolled sources in the Permian Basin. Our DI method also introduces new features to the application of the divergence integral adapted for our airborne mission, as detailed later in this paper.

2 Data and Methods

2.1 MethaneAIR Research Flights

In addition to the controlled-volume experiments, we report here MethaneAIR research flights focused on the Permian Basin, straddling the Texas-New Mexico border, and the nearby Midland-Odessa areas where O&G activities are abundant (see Table 1). ~~All the~~ Methane from nearby sources can confound the identification and quantification of emissions from a particular source, and data from these flights helps us to understand the impact of adjacent sources on our quantification of emissions from point sources in the real world. All the MethaneAIR flights originated from the Rocky Mountain Metropolitan Airport in Broomfield, CO, using the NSF/NCAR GV HIAPER Aircraft (UCAR/NCAR - Earth Observing Laboratory, 2005). Research flights RF04 and RF05 targeted the volume-controlled releases, plus validation to an EM27/SUN solar viewing spectrometer in East Colorado. The plane repeated overflights of the release site from around 12960 m altitude, at approximately 20-minute intervals, for up to 6 hours while the crew on the ground manipulated the release rates. Methane emitted from nearby O&G activity interfered with the observations on at least 3 overpasses. ~~We obtained~~ The first three research fights (RF01 - RF03) were successful engineering fights yet omitted from the paper since our main purposes were for instrument functionality confirmation rather than methane data collection. Here, we show the results of 9 and 12 cloud- and interference-free images for RF04 and RF05, respectively. The flight tracks can be found in the supplement (FigureS32, FigureS33)

Flights RF06 and RF07 aimed to map the Delaware sub-basin of the Permian oil and gas field. The strategy for mapping the areas of interest is to create repeated tracks over the target areas with partial overlaps (i.e., area mapping). The overlapped segments act as buffers of missing data, allowing us to observe the changes in the column-averaged dry-air mole fraction of methane (XCH_4) over time.

Table 1. Summary of research flights of interest during the MethaneAIR campaign over the summer of 2021

Research Flights	Dates	Targets
RF04	30 July 2021	Blinded-volume controlled releases & EM27/SUN
RF05	3 August 2021	Blinded-volume controlled releases & EM27/SUN
RF06	6 August 2021	Delaware Sub-basin
RF07	9 August 2021	Delaware & Midland Sub-basins

During the survey of the Delaware Sub-basin of the Permian Basin flight (RF06), MethaneAIR covered more than 9,000 square kilometers in 2.2 hours. An example of what MethaneAIR observed from a source of methane is shown in Figure 1 (a). This source was measured during RF06 and identified as the MiVida Gas Processing Plant in Barstow, TX. When the flight tracks are overlapped, we can combine all the data segments and create a mosaic map of the entire area of interest as shown in Figure 1 (b).

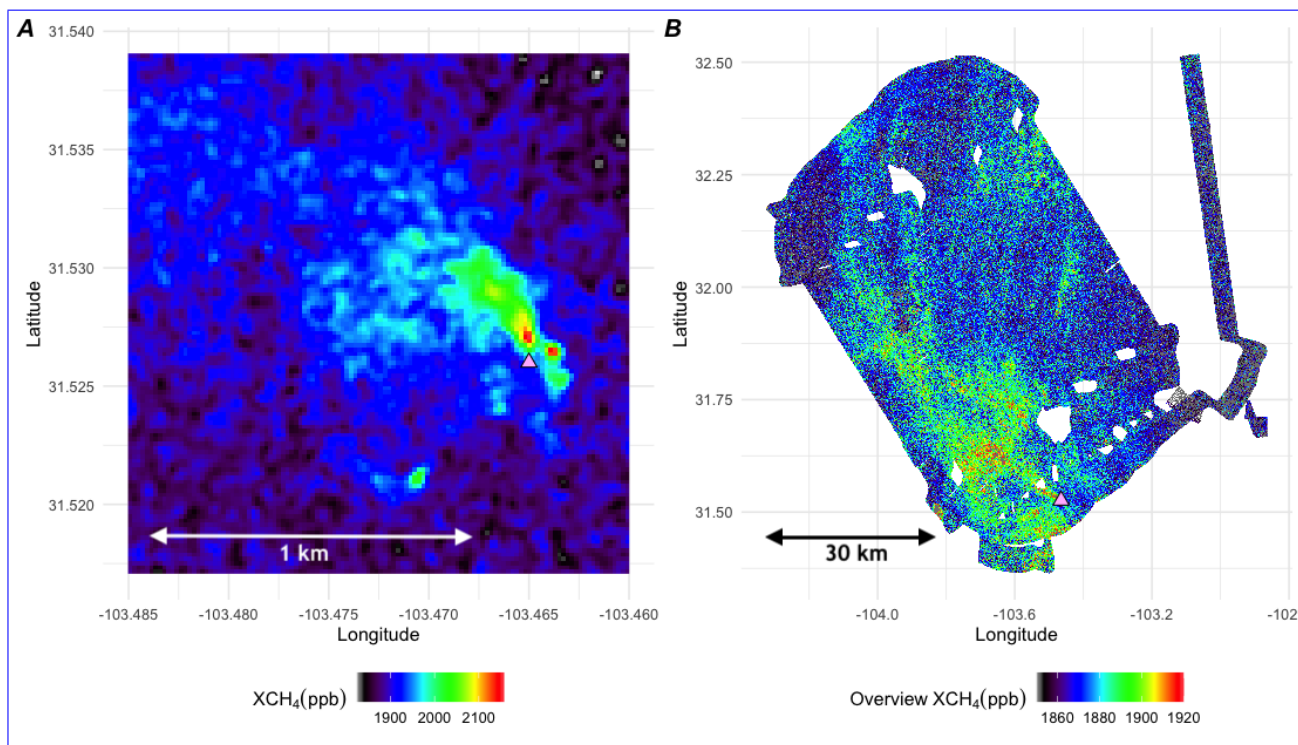



Figure 1. Examples of the MethaneAIR data on 6 August 2021 (a) MiVida Gas Processing Plant (pink triangle) from RF06 segment 10 (approximately 2km x 2km image) and (b) mosaic image (75 km x 120 km) of the entire RF06 scene from all the segments including the segment shown in Figure 1. The prominent gap areas in this picture were not imaged during the overlapping turns of the pattern. Full flight track information can be found in S5.

110 In 2022, we conducted another single-blind volume-controlled experiment over two research flights near Pheonix, AZ. Only RF01E and RF03E research flights that include the experiment are shown here. The flight tracks can be found in the supplement (FigureS34, FigureS35). We obtained results of 11 and 13 cloud- and interference-free images for RF01E and RF03E, respectively.

Table 2. Summary of research flights of interest during the MethaneAIR campaign over the fall of 2022

Research Flights	Dates	Targets
RF01E	25 October 2022	Blinded-volume controlled releases & EM27/SUN
RF03E	29 October 2022	Blinded-volume controlled releases & EM27/SUN

2.2 MethaneAIR Level 3 Products

115 From our cruise altitude of 12960 m above ground ((13850 above sea level), the across track distance between pixel centers is 5 m and swath of about 4.5 km or 863 pixels. The along-track sampling has centers ~25 m apart. The MethaneAIR XCH_4

data were retrieved from the 1.65-micron band, with CO_2 as the retrieval proxy for the optical path, using the Smithsonian PLanetary ATmosphere interface to VLIDORT (SPLAT-VLIDORT) radiative transfer code (Chan Miller et al., 2022).(Chan Miller et al., 2023) Our multilayer methane retrieval algorithm is based on the CO_2 proxy method. We use true methane columns comprising the paths from space to the ground and the ground to the aircraft ($\approx$ 12 km). We account for the influence that topographic variations will have due to the profiles of methane and carbon dioxide using the averaging kernel for our sensor. The tropopause height variations may influence the methane columns due to less methane in the stratosphere. The influence may play a role at some MethaneAIR altitudes and will certainly affect the MethaneSAT data since the satellite will be higher up in the atmosphere. Since most excess methane is within the boundary layer (below the aircraft), we assume the averaging kernel for anomalies applies for the lowest kilometers. The averaging kernel is slightly larger near the surface than in the upper atmosphere. Based on the engineering flights, the albedo dependence for MethaneAIR data is lower than anticipated, given that during the campaign, observations were often over regions blanketed by haze from long-range transport of smoke from fires in the Western United States and Canada. The size distribution may be small enough for the aerosol optical depth to be insignificant at 1600 nm since a large fraction may be in drier air in the free troposphere. The sign and magnitude of albedo bias will be scene-dependent, determined by the prior profile bias and additional light-path modifications induced by aerosol scattering. The detailed discussion can be found in Chan Miller et al. in prep.

The XCH_4 data were later projected onto a 10 m $\times$ 10 m grid; in our report to the blinded experiment, we used a nearest-neighbor gridding method that accounts for sensor oversampling on the ground (Appendix S1.4), whereas in our post-unblinding ("best") rendering, we used the "snowflake" gridding approach of Sun et al. (Sun et al., 2018). The gridded images were denoised using a 2-dimensional mass conserving Gaussian filter with a half-width at half maximum of 0.85 pixels, yielding denoised XCH_4 10m $\times$ 10m gridded data with the noise of 40 ppb (1σ) for the blinded report and 15-20 ppb (1σ) for the best rendering. The spatial decorrelation length for the image is $\sim$70m.

2.3 Methane Point Source Emission Quantification

We developed two main methane point source quantification approaches:, mIME and DI, both adapted from the literature for application to MethaneAIR data. For the purpose of testing the To test the accuracy of our quantification methods, independent of a plume-finding exercise, we assumed that the source location was already identified. We note that, by assuming that source locations were identified, we have a lower chance of having false positives.

2.3.1 Large-Eddy Simulation Based Approaches

We used the Weather Research Forecast Model driven by High-Resolution Rapid Refresh (HRRR) meteorological data in the Large-Eddy Simulation mode (WRF-LES-HRRR) to simulate plumes at the source point. The simulated plumes exhibit stochastic behavior in response to eddy-scale winds associated with boundary layer turbulence, providing quantitative information on the uncertainty due to these processes.

The WRF-LES code we used was first developed to simulate CO_2 transport in the US Upper Midwest and Indianapolis areas (Lauvaux et al., 2012; Gaudet et al., 2017). Our WRF-LES model can take GRIB2 files of the meteorological data of

150 interest (e.g., HRRR, European Centre for Medium-Range Weather Forecasts, and Global Forecast System data) and static geographical data as inputs. For this study, we used HRRR winds to drive the WRF-LES combined with the highest resolution static geographical data available on the official WRF website (Blaylock et al., 2017; Skamarock et al., 2008). ~~We~~ Given the knowledge of point source locations, we nested (one-way) the domains starting from the native HRRR resolution of 3 km × 3 km to 111 m × 111 m grid cells by nesting the domains three times at 3:1 ratio each time. The innermost domain is 103 pixels

155 by 103 pixels or 11.44 km by 11.44 km. There are 49 vertical layers, with the highest resolution near the surface and coarser towards the top of the atmosphere. We used the Noah Land-Surface Model scheme for the WRF surface option, 1.5 order TKE closure (3D) for the eddy coefficient option of the innermost domain, and Mellor–Yamada– Nakanishi–Niino (MYNN)-Eddy Diffusivity-Mass Flux (EDMF) scheme (Nakanishi and Niino, 2006, 2009). The closure in the boundary layer limits effective eddy resolution to 50 m – 100 m; by limiting our grid to 111 m, the run time for the model is tractable, and the simulation of

160 the plumes is realistic on a spatial scale approaching MethaneAIR resolution. We ran the simulations for the days of interest, starting 5 hours before the first observation on each day and ending half an hour after the last observation of interest.

The center of the innermost domain was set to align with the emission source location. When multiple sources are clustered in close proximity, within plus or minus 1/3 the size of the domain from the center (approximately 3.5 km for most of our innermost domains), those sources can be placed in the same model run without expanding the size of the domain. If the

165 sources spread out more than 1/3 of the domain size, we can expand the domain sizes accordingly with additional computation time. Our WRF-LES-HRRR writes innermost domain outputs every minute. The outputs include all the ~~parameters of the model~~ model parameters and a specific ~~concentration of methane~~ methane concentration for every point in the domain.

We identified and isolated the plume from each point source by defining a threshold above inflow concentration, 1.5 standard deviations of the inflow values, above the median value of the inflow, into the source region (see Figure 2B for an example of

170 an isolated plume and Figure S10 for an example of emission rates as a function of thresholds), setting values below that values to NA. Since we have full knowledge of the emission rate underlying the plume from each LES scene, we could, in principle, determine the "effective wind speed" (U_{eff}, the turnover rate for the mass in the plume; Varon et al. (2018)) that would give the best fit to the observed plume shape for each encounter, and then estimate the emission rate from the ratio of the mass of the observed plume to the mass of the LES plumes times the nominal emission rate given to the LES model. We used this method,

175 called the Ratio Method (Appendix S1.3), in Irakulis-Loitxate et al. (2021) where the resolution of the plume images was close to the LES resolution. However, here the LES spatial resolution is coarser than MethaneAIR resolution, so we modified the Integrated Mass Enhancement (IME) method of Varon et al. (2018) to adapt to our conditions, as described below.

Modified Integrated Mass Enhancement

The concept of *IME* ~~is to compute the~~ emission started with the computation of the total mass enhancement in a defined

180 plume~~and multiply the~~ . The *IME* is the integrated mass enhancement of the plume and defined as

$$IME = \sum_{j=1}^{N} \Delta\Omega_j A_j. \tag{1}$$

where $\Delta\Omega_j$ is the mass enhancement ~~by the turnover rate for the mass in the plume (U_{eff}/L). The mass enhancement~~ per pixel~~is~~, defined as the total column of methane enhancement relative to the inflow concentration ~~per area ($\Delta\Omega$) times~~ and A_j is the area of each pixel ~~. The *IME* equation is~~ j. To obtain the emission rate, we introduce the following equation

$$Q = \frac{U_{\text{eff}}}{L} IME, \tag{2}$$

where Q is the emission (kg/hr), L is the scale length defined as the $\sqrt{A}$, and A is the area of the plume (m^2). We multiply the *IME* by the turnover rate (U_{eff}/L). The effective wind speed U_{eff} is a function of $\log(U_{10}) + 0.6$, where U_{10} is the wind speed at the 10m height. ~~The *IME* is the integrated mass enhancement of the plume and defined as~~

$$IME = \sum_{j=1}^{N} \Delta\Omega_j A_j.$$

~~The concept of IME is to compute the total mass enhancement in a defined plume and multiply by the turnover rate for the mass in the plume (U_{eff}/L). The mass enhancement per pixel is the total column of methane enhancement relative to the inflow concentration per area ($\Delta\Omega$) times the area of each pixel. The~~ The Varon et al. (2018) IME method used U_{10} from operational meteorological products, whereas our modified approach (mIME) uses the 10 m root-mean-square wind from each LES realization specifically run for the case of interest, averaged over $\pm$ 15 minutes from the sampling time.

The relationship between U_{10} and U_{eff} is given by a set of empirical coefficients (Appendix S1.1; (Varon et al., 2018)), reflecting the similarity of the root-mean-square winds in our LES runs to the winds from the large ensemble of (idealized) LES simulations in Varon et al. (2018). Since our LES winds simulate the eddy-scale component applicable to each plume, our simulations give us an estimate of the variability of the IME at the time of each observation, providing the uncertainty for each event.

2.3.2 Divergence Integral Method

The DI method is based on Gauss' divergence theorem applied to surfaces enclosing the source. The surface integral is estimated based on XCH$_4$ measurements along a rectangle surrounding the source ~~by combining Eq. ??~~ following Eq. S5 and Eq. ~~S5~~S6:

$$\Phi_{surf} = \sum^{\text{around rect}} (XCH_{4_i} - \langle XCH_4 \rangle_{rect.}) \cdot n_{\text{column}} \cdot M_{CH_4} \cdot v_{\text{perpendicular}} \cdot \Delta l + \left\langle \frac{\partial m}{\partial t} \right\rangle \tag{3}$$

where Φ_{surf} is the flux (kg/hr) into the enclosed volume from the land surface, XCH_{4_i} is an individual measurement along the rectangle, $\langle XCH_4 \rangle_{rect.}$ is the mean of all measurements along the rectangle, Δl is the distance between successive XCH_4 measurements, n_{column} is the moles of air in the column based on the surface pressure from HRRR, M_{CH_4} is the molar mass of methane, $v_{\text{perpendicular}}$ is the wind speed perpendicular to the sides of the rectangle, and $\left\langle \frac{\partial m}{\partial t} \right\rangle$ is the rate of change of

total mass in the enclosed volume (∼0 for plume-size volumes). The wind speed was obtained from HRRR but was rotated to match the wind direction of the observed plume using the difference between the major axis of the moment of inertia and the HRRR wind direction (see Appendix S1.2). Errors in wind direction and wind speed from HRRR will affect the DI method's calculated emissions. We used the wind direction information in the plume image to reduce this potentially large source of error.

The flux was calculated for a series of rectangles ("expanding boxes") with the distance from the source to the downwind edge of the rectangle ranging from approximately 100 m to the length of the observed plume (typically 300 - 2000 m). To fully take advantage of the information in the MethaneAIR image, the size of the rectangle was increased sequentially by one pixel (10 m) in each direction, with the exception of the upwind side, which was increased by a pixel for every fourth rectangle to avoid catching other sources yet still average over the background (Figure 2c).

Figure 2d shows the calculated fluxes around the unlit flare near the controlled release on 3 August 2021 for individual rectangles as a function of the distance from the source to the downwind edge of the rectangle. Boundary layer eddies break up the plume structure, leading to a buildup of methane concentration in some areas and depletion in others. Since these eddies are not resolved by HRRR, we see oscillations of the computed flux values as a function of distance, on the scale of the associated structures. However, by averaging the flux over the range of distances, we arrive at a robust flux estimate for the source that averages over the eddy-scale variability of the winds. The estimated flux for the individual source is the average flux for all rectangles crossing the plume, in this case, rectangles with edge distances from 100–800 m. (*Note*: The blinded DI emission estimate initially reported to Stanford used the alternate name "Gaussian Integral" for this method).

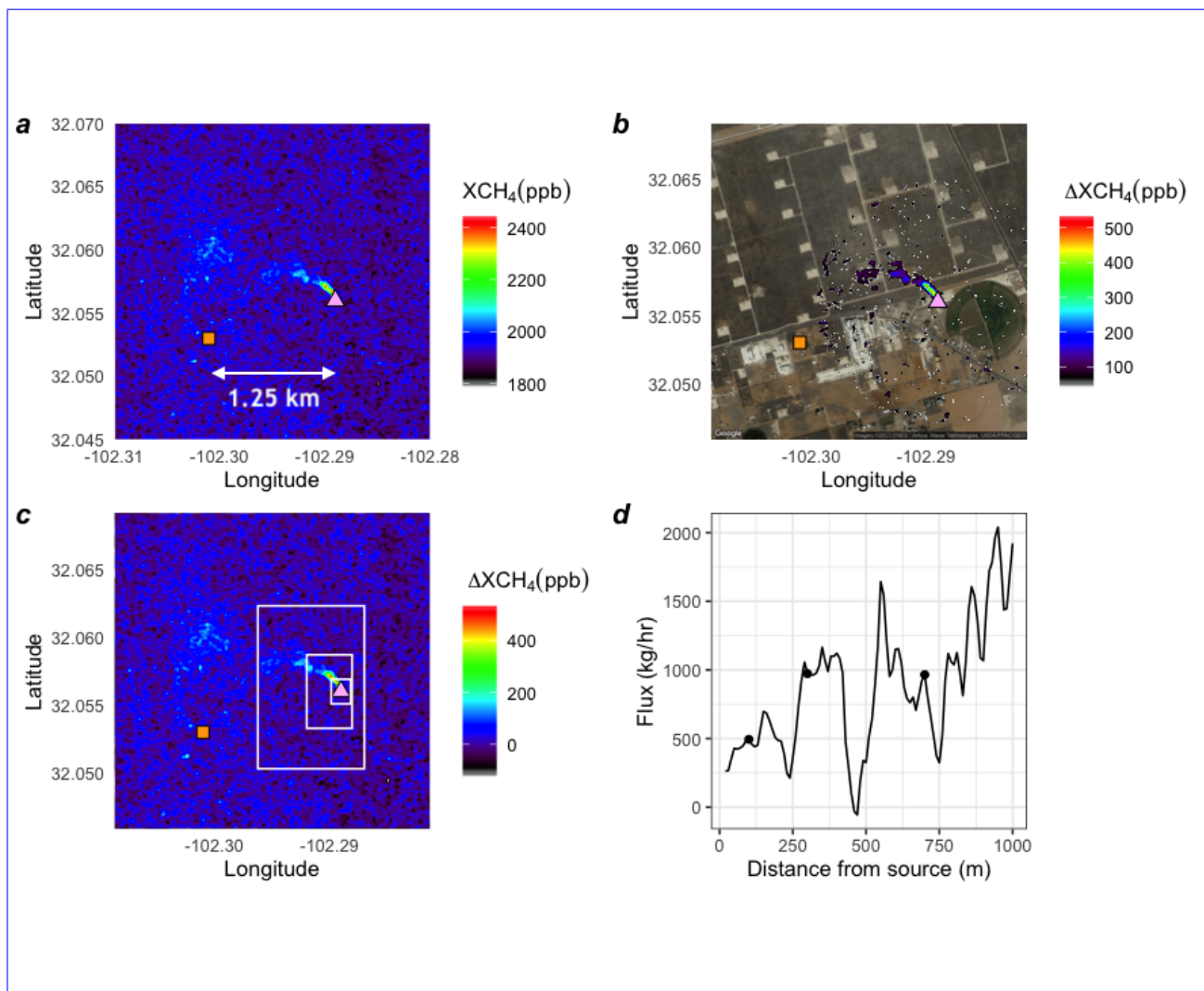



Figure 2. a: MethaneAIR image of a blinded volume-controlled release (orange square) and an unlit flare (pink triangle) observed on 3 August 2021. **b**: A sample of the isolated unlit flare plume using the mIME method. The isolated plume is superimposed on a ©Google Map 2022 satellite layer. **c**: Three example rectangles are used to calculate the flux divergence are shown. **d**: Calculated flux divergence as a function of distance from the source to the downwind edge of the rectangle. Circles indicate the position of the example rectangles are shown in Figure 2c. The fluctuations of the apparent flux with distance from the source reflect the influence of eddy scale motion as well as contributions of excess methane from nearby sources. The influence of eddy-scale motions is evident in the oscillation at approximately 250 m intervals, the apparent length scale for eddies at this overpass. The surface flux is estimated by averaging the flux divergence over several eddy scales to average out $\partial m/\partial t$. In this example, we only averaged the DI from 80 – 700m to avoid influence from other sources nearby that increase the DI beyond 700m from the source.

2.4 Wind Validation

Since the accuracy of the winds directly affects the emission estimates, we examined other sources of wind data to assess

230 the associated uncertainty. We compared our modeled WRF-LES-HRRR winds to wind observations from nearby airports as well as ground-based wind measurements from an instrument at the controlled release site. The comparisons suggest that WRF-LES-HRRR winds are comparable to ground-based wind measurements near the site while the larger scale or nearby airport winds generated larger errors.

2.4.1 Ground Wind Validation

235 Wind speeds at 10 meters above ground level were measured by a Gill Instruments WindSonic 60 2-dimensional ultrasonic anemometer. The anemometer was placed on a 4 m pole-mounted tripod, secured to the steel flooring of a JLG 400 s telescopic boom lift bucket. The boom lift bucket is designed for a vertical lift while keeping the platform level, and fine adjustments to its height could be made. The anemometer was placed roughly 40 m S-SE from the releasing stack. The data were withheld from the MethaneAIR team before the first data comparison. We later used measured winds from this site to check our estimations

240 from WRF-LES-HRRR.

2.4.2 Automated Surface Observing Systems Wind Validation

Automated Surface Observing Systems (ASOS) provide meteorological measurements at airports in the United States, including wind surface observations. The data are reported continuously every minute and released to the public within two weeks (US Department of Commerce). ASOS winds can be used as quality checks for wind products used in MethaneSAT

245 or MethaneAIR emission estimates when ground-based measurements at the source are not available. The ASOS resolution is 1 kt or 0.5 meter per second, limiting its utility under low wind conditions.

We compared the ASOS data from nearby airports to the average WRF-LES-HRRR winds near the source. The time series of the winds from different products are shown in Figure S8. The purple crosses represent ASOS winds from the Midland International Air & Space Port (MAF). Because the locations were slightly different, we only look at the variation of the winds

250 within the window of interest. When the variations of the ASOS winds and the WRF-LES-HRRR winds are from the same distribution (two-sided t-test p-value is greater than 0.05), the WRF-LES-HRRR winds are trusted. For the days of blinded-volume controlled release experiments, all the p-values were greater than 0.05 and the LES winds were used in all cases.

2.5 Error Analysis

Since resampling the same view of a plume multiple times is not feasible due to the changing meteorological conditions and

255 flow rates, we used the Monte Carlo simulation principle to calculate the confidence intervals of the mIME estimates. ~~For each plume~~Based on the original observation, we generated a ~~set of one~~ thousand synthetic observations~~from the observation. We resampled~~. To simulate a synthetic observation, we sampled XCH_4 pixels of the ~~originally~~ observed inflow with ~~a replacement for each synthetic observation~~replacement n times, where n is the number of the original inflow pixels. The new resampled

inflow mean and standard deviation values were used as a new background and a threshold. After applying the new threshold
to the field of XCH_4, we constrained the extent of the new plume using the original area. A new mass was calculated from the
new plume. For the winds, we randomized the mean inflow winds by sampling from seven LES snapshots, taken at 5-minute
intervals within 30 minutes of the retrieval. We calculated the mIME estimate for each resampled set of values. Finally, we
calculated the confidence interval (the 2.5[th] and 97.5[th] percentile of the emission estimates) from the 1000 resampling trials.
Our mIME confidence intervals, therefore, include the variances from the LES winds, instrument, and measurement errors for
this particular set of MethaneAIR determination of emission rates.

For the DI method, we used a t-test on the set of fluxes obtained for boxes of different sizes expanding in the downwind
direction from the source to estimate the 95% confidence interval of the fluxes. Schneising et al. (2020) used a similar approach
but only varied the position of the downwind side of the rectangle, leaving the other sides fixed. By changing the position of
all sides of the rectangle, we average over different clean segments and reduce errors based on concentration fluctuations
outside of the plume (unlike our application, Schneising et al. (2020) were able to average over a longer time series enabling
them to average out this uncertainty, but we have only one snapshot). We use this method on a much different length scale,
as Schneising et al. calculated basin-level emissions, on the scale of 400 km, compared to our goal of quantifying individual
emitters with plumes spanning 0.4 to 10 km.

2.6 Targeted Emissions Sources

During the MethaneAIR campaign in the Permian over the summer of 2021, we applied mIME and DI approaches to obtain
the emission rates, validated the winds, and estimated the errors for the blinded volume-controlled releases of methane and
other uncontrolled methane sources in the area.

2.6.1 Blinded-Volume Controlled Release Experiment

Controlled release testing was conducted to evaluate the quantitative performance of MethaneAIR. We conducted repeated
passes over the metered methane point source during controlled release intervals. The MethaneAIR team was aware of the
release location but not of the release rate of the methane source nor whether any release was happening at a given over-
pass. Stanford University organized the controlled release experimental campaign (32.053 °N, ~~102.301~~ 102.301 °W) near
Gardendale, Texas, in July-August 2021, sampled on two days of the campaign. A suite of other teams in controlled release
testing, including multiple aerial teams ~~(Rutherford et al. in prep)~~ (Rutherford et al., 2023) and satellite teams ~~(Sherwin et al~~
~~2022)~~(Sherwin et al., 2023). MethaneAIR participated in the campaign on July 30 and August 3, 2021.

MethaneAIR was in an earlier stage of development than other tested systems, having never obtained any scientific data
prior to the sampling controlled release of methane (limited engineering data had been obtained in 2019). Due to this novelty, a
slightly modified testing agreement was created with a collaborative publication plan in lieu of independent publication results
by the Stanford team, as was performed for other teams. Other protocol changes include: (1) MethaneAIR submitted three
sets of estimates using different algorithms and selected which was considered to be best, before unblinding (2) MethaneAIR
requested that information about wind direction and the presence or absence of meaningful wind speeds be provided, which

Stanford sent to Harvard during the experiment seven times. This information was intended solely to inform the aircraft's flight trajectory, which was not in fact modified from the filed flight plan during the testing.

The limitations of this experiment include (1) a small range and low flow rates of the controlled release values (all less than 1000 kg/hr) ~~and~~ compared to larger than 1000 kg/hr releases observed by other participants, (2) the release sites were shifted between 3 m and 5.25 m altitude and (3) the number of blinded-volume controlled releases measured was limited to 21 points on two days. The experiment was, however, fully blinded — the metered emission values were not revealed to the MethaneAIR team until after initial emission estimates from MethaneAIR using the mIME and DI methods were reported to the Stanford team, including all overpasses where the IME method detected an XCH_4 plume above a 1.5 standard deviation threshold within one kilometer from the releasing site. An unplanned limitation was imposed by significant methane emissions from sources close to the releasing point, including an unlit flare with a high emission rate located at 32.056°N, 102.288°W, ~1 km to the east. This suggests that sources within 1 km may impact the ability to quantify emissions. Emissions from this source were detected on all passes of the blinded-volume controlled release experiment. MethaneAIR has higher sensitivity and a large pixel size (10m x 10m or 20m x 20m, depending on the gridding choices) than other airborne sensors. The methane from this and other extraneous emission sources sometimes interfered with MethaneAIR observations, especially for the DI.

After unblinding of the 2021 experiment, we observed another set of single-blind volume controlled releases near Pheonix, AZ (32.821 °N, 111.786 °W) on 25 & 29 October 2022. Stanford's 2022 instruments have lower uncertainty than those used in the 2021 experiment (El Abbadi et al., 2023). With better precision in the instruments, the Stanford team reported the 2022 emission rates as 60-second means instead of 90-second means in 2021.

2.6.2 Emission Rates for Miscellaneous Methane Sources in the Delaware Basin

Many miscellaneous methane sources were detected during the MethaneAIR research flights RF06 and RF07. We compared the two emission estimates, mIME and DI, and report here both values. We manually identified the scenes with high ~~concentrations of methane~~ methane concentrations and co-located with any O&G infrastructures or reported leaks. The sources from RF06 and RF07 enabled us to estimate methane emissions from 9 overpasses of the MiVida gas plant (31.524°N, 103.467°W, Bartow, TX) over two days. We discovered and quantified a very large pipeline leak in New Mexico on RF07.

3 Results and Discussion

3.1 Blinded Volume-Controlled Release Experiment

For two testing days, MethaneAIR collected 21 data points. We included emission estimates based on the mIME, the DI, and the Ratio methods presented in Table S3 as the original results submitted to Stanford by MethaneAIR on 1 February 2022 (~~prior to~~ before unblinding the metered release volumes). Based on the decision tree workflow in Appendix S3.4 using both mIME and DI results, a set of blinded best-estimate emissions was submitted to Stanford as the original submission. In this original blinded submission, one emission estimate of the 21 was made using the mean of DI and mIME, and the rest were

made using mIME only. After the initial report of the blinded best-estimate data, we decided to use only the mIME method for the consistency of all the data points. Emission rates were not modified post-unblinding. However, some edge cases were flagged via the decision tree. Both blinded best estimate and post-unblinding results are presented in Table S1. Only post-unblinding mIME results are presented in Figure 3. For references, the post-unblinding DI results are presented in Table S2. The post-unblinding mIME and DI results are also presented in Figure S1 and Figure S2 as bar plots representing the emissions and ~~the decision-tree flags~~color-coded following the decision tree. In addition, the post-unblinding DI estimates are presented in Figure S4 as an analogous plot of Figure 3 similar to the blinded best-estimate results that are presented in Figure S3.

After the Stanford team provided the true emission rates from the experiment, we designed an algorithm that could reliably reject mIME estimates that were false positives or below the detection threshold for a point source. We also developed a more comprehensive bootstrapping approach to account for uncertainty from background concentrations and thresholding in addition to the winds. In the second set of results, this filtering algorithm designated "below detection limit" six emission passes from the dataset; the quantified emissions were slightly changed after unblinding due to the new bootstrapping results. One flagged data point: MethaneAIR estimated the emissions to be above 200 kg/hr when the metered emission is below 200 kg/hr; this observation suggested that, when emissions are below the detection limit, the estimated emissions are uncertain and can be overestimated. These flagged and no-detection data points were not removed before we applied the OLS or York regressions. Based on this comparison, we determined that the detection limit for the MethaneAIR 1x1 retrievals using the mIME method was approximately 200 kg/hr. This value is preliminary because it is based on a small sample size, for two days, in contaminated areas; future work is needed to fully characterize the detection limit (such as using methods presented in Conrad et al. (2022)).

Figure 3 shows methane flux rates determined by the MethaneAIR mIME method compared to the controlled release rates from the Stanford team from RF04 and RF05 flights on 30 July and 3 August 2021, respectively (all the plumes identified and used here are shown in Appendix S4), including the minor adjustments after unblinding and designating with red circles the false positives in the blinded submission. The red square designates the false negative in the blinded submission. The York regression slope (Appendix S1.5, MethaneAIR (y-axis) vs blinded-volume controlled releases (x-axis), is 0.96 ($\pm$ 0.12, 95% confidence interval), using error estimates from our bootstrap (y-axis, mIME emission rates) and from the Stanford team (x-axis, reported metered flows). The intercept is 88 kg/hr ($\pm$ 27 kg/hr). Since the uncertainties in the controlled release rates are not negligible, the York slope provides the maximum likelihood estimate of the relationship between controlled release and emissions obtained from analyzing MethaneAIR data. We compare this fit to the ordinary least square (OLS) regression results, often used to analyze this type of experiment, where the slope is 0.85 ($\pm$ 0.13, 95% confidence interval), R of 0.83, and the intercept is 113 kg/hr ($\pm$ 45 kg/hr). The limitations of this experiment include (1) a small range of blinded-volume controlled release values (all less than 1000 kg/hr) and (2) a limited number of blinded volume-controlled releases measured (N = 21). The excellent agreement between the fluxes calculated by applying the mIME method to MethaneAIR data and the true controlled release rates provide strong support for the validity of our XCH_4 retrievals and flux calculations, subject to those limitations.

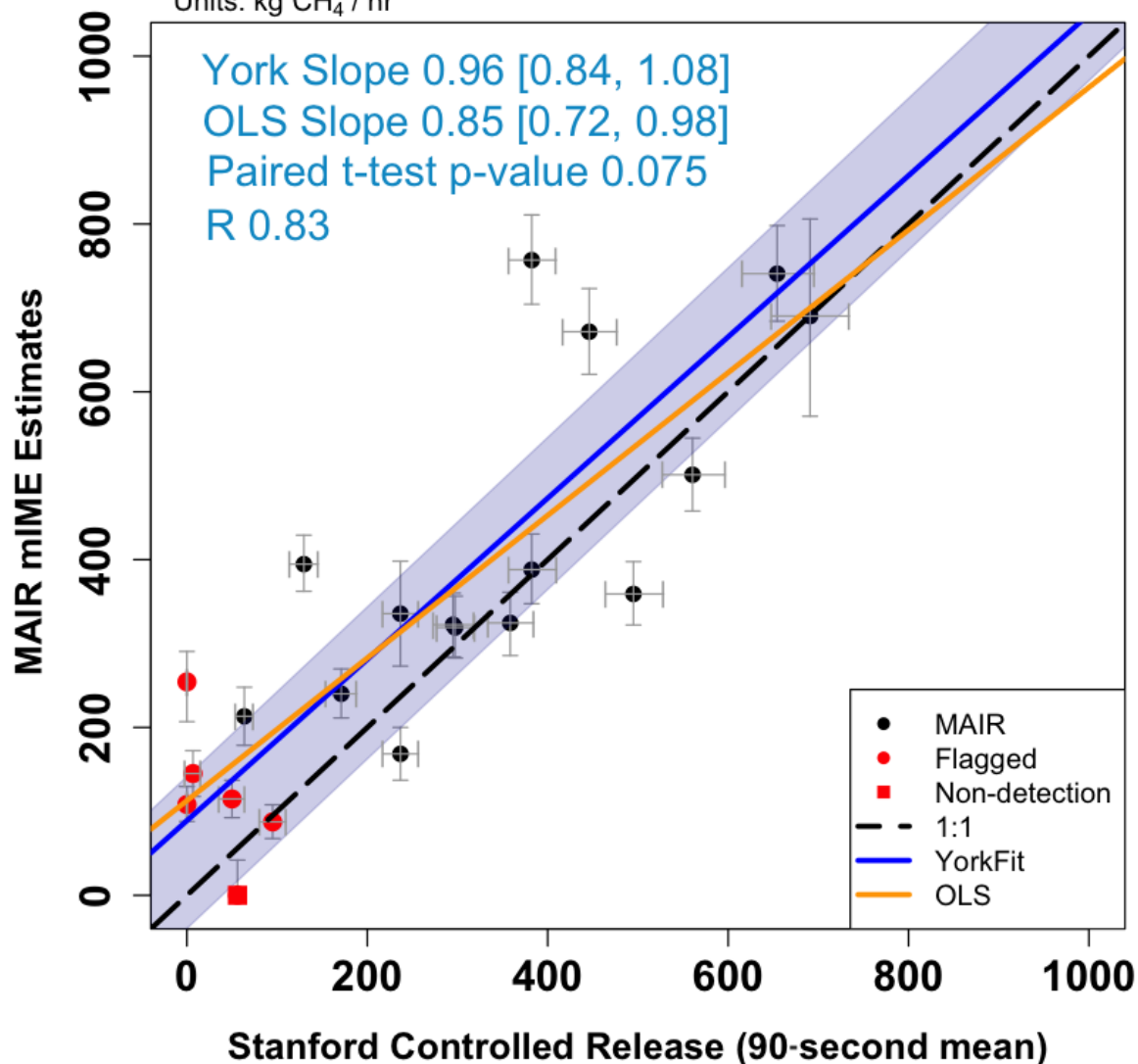



Figure 3. Volume-controlled release experiment results after unblinding. The results presented in this figure are based entirely on the mIME method. The black circles represent the post-unblinding estimates plotting against the reported metered emissions. The red square represents no detection. The red circles represent data flagged as below the detection limit (S11). The blue solid line is the post-unblinding York fit. The blue shaded area represents a 95% confidence interval of the York fit from the resamples. The orange is the Ordinary Least Squares (OLS) fit. The p-value of 0.075 from the paired t-test between the estimated emissions and Stanford emissions suggests that we cannot reject the null hypothesis that the population mean of the differences is 0 with a 95% confidence interval.

We conducted paired sample t-tests to determine whether the mean difference between the reported emissions from the Stanford team and the estimated emissions is different than zero. Our null hypothesis is that there is no difference between the mean observed and estimated emissions. The test was performed three times for three sets of estimated emissions based on three different wind products: WRF-LES-HRRR winds, HRRR winds, and measured winds at the release site (See Figure S9 for the comparison). The p-values are 0.17, 0.041, and 0.31, respectively. We cannot reject the null hypothesis of no true mean difference between the reported emissions and the estimated emissions derived from WRF-LES-HRR or observed winds. However, the 0.0408 p-value for HRRR winds suggests that there is likely to be a true mean difference between the reported emissions and the estimated emissions. We infer that the WRF-LES simulated winds improve the mIME method compared to simply using HRRR winds in the IME function. Within our limited data, the mIME method works equally well using WRF-LES-HRRR winds or measured local winds.

For two testing days in 2022, MethaneAIR collected 24 data points. We included emission estimates based on both the mIME, the DI, and the best estimate (the average between the two methods) as results submitted to Stanford by MethaneAIR on 22 March 2023 (prior to unblinding the metered release volumes).

Figure 4 shows methane flux rates determined by the average of the MethaneAIR mIME and DI methods compared to the controlled release rates from the Stanford team from RF01E and RF03E flights on 25 and 29 October 2022, respectively, including the low-quality estimates designating with red circles in the blinded submission. As a result, the confidence intervals are much larger than the confidence interval in 2021, which was exclusively determined by the mIME method.

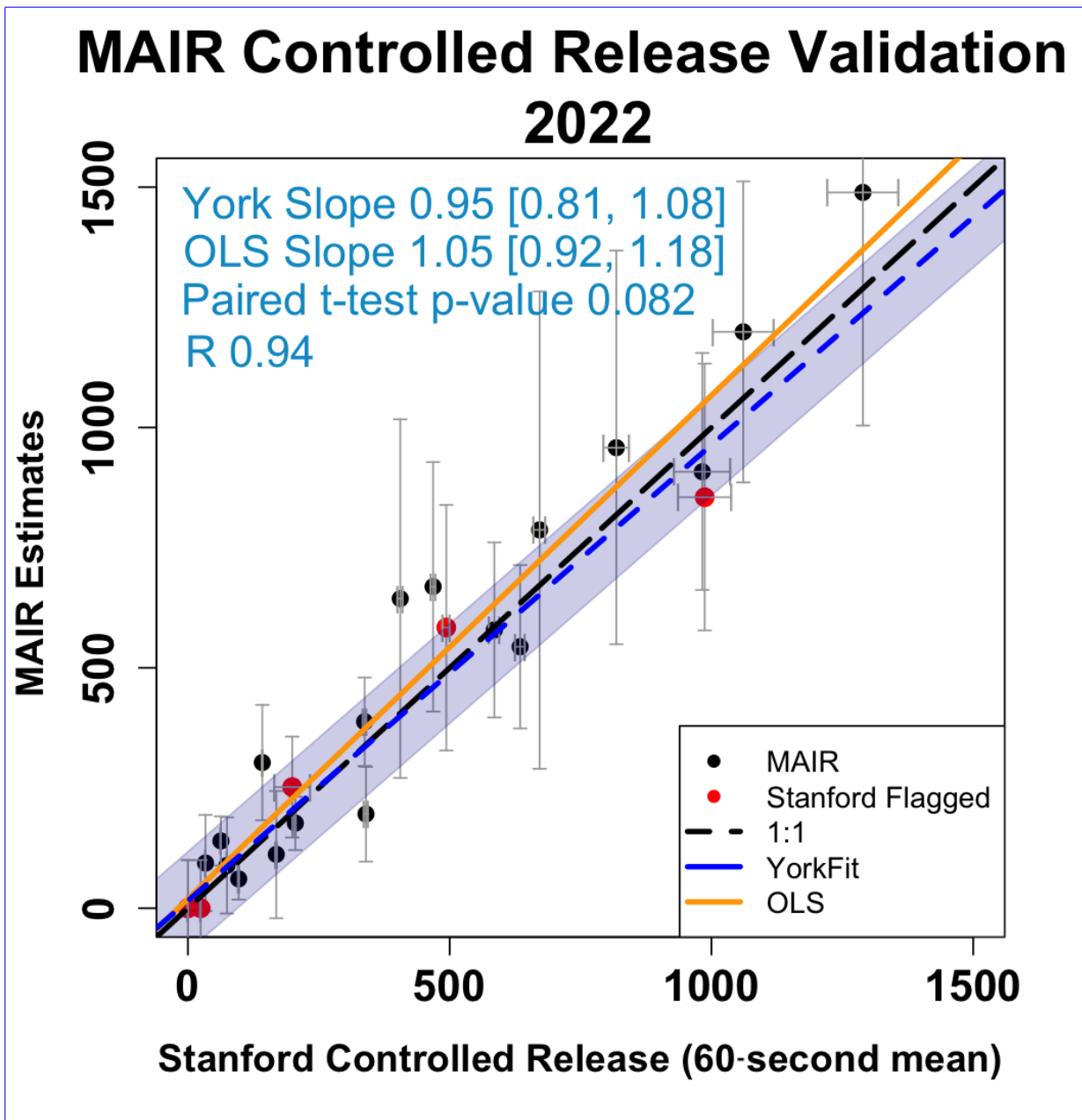



Figure 4. Volume-controlled release experiment results after unblinding. The results presented in this figure are based on the average between the mIME and the DI method. The black circles represent the post-unblinding estimates plotting against the reported metered emissions. The red circles represent data flagged as low-quality data points by the Stanford team. None of these data points are flagged by MethaneAIR. The blue solid line is the post-unblinding York fit. The blue shaded area represents a 95% confidence interval of the York fit from the resamples. The orange is the Ordinary Least Squares (OLS) fit. The p-value of 0.082 from the paired t-test between the estimated emissions and Stanford emissions suggests that we cannot reject the null hypothesis that the population mean of the differences is 0 with a 95% confidence interval.

3.2 Miscellaneous Methane Sources

3.2.1 Unlit Flare

375   MethaneAIR observed the unlit flare near the release site 21 times with repeated overpasses on 30 July and 3 August 2021. As shown in Figure S6, the flare emission rates ranged from 500 kg/hr to 2000 kg/hr as calculated by the mIME and DI methods. Based on this comparison and the comparison with the controlled releases, we estimated the detection limit of the DI method to be approximately 500 kg/hr.

3.2.2 MiVida Gas Processing Plant

380   We quantified emissions from a large consistent emitter, the MiVida Gas Processing Plant, in Barstow, TX, 9 times with repeated overpasses on 6 and 9 August 2021. Typical emission rates ~~approximately~~ were 2000 kg/hr with 30% uncertainty estimate based on the controlled releases experiment results (Figure S7). Several emission points could be identified within the boundaries of this plant.

3.2.3 The Pipeline Leak in New Mexico

385   We discovered a large pipeline leak at 103.697°W, 32.366°N in New Mexico. MethaneAIR detected the leak on 9 August 2021, not present on the earlier flight on 6 August 2021. Using the mIME and DI methods, we estimate the emission rate to be approximately 5000 kg/hr with 30% uncertainty estimate based on the controlled releases experiment results (Figure S7).

We attributed the leak to a gathering pipeline emission event at the location, confirmed by emission incident reports to the New Mexico Oil Conservation Division on August 26, 2021, which identified the cause as a ~~rapture~~ rupture at a weld along
390   a gathering pipeline segment (OCD, 2021). Based on this report, we estimated a methane emission rate of 8200 kg/h over an 18-hour reporting period, assuming 80% methane content in the gathered natural gas. With simple extrapolation of the lower estimate of 5000 kg/hr from the flight on August 9 until the leak was fixed on August 24, the total methane emission could be more than 1.8×10^6 kg (1,800 metric tons) over the 15 days.

3.3 Confidence Levels of the Emission Estimates

395   Beyond the confidence flags from the decision tree (Appendix S3.4) and attempts to understand the methane ~~contaminants~~ contamination of the scenes of interest by methane from nearby sources (Appendix S3.5), we also evaluated the validity of our ~~instrument~~ emission estimates by plotting the methane emission rates from both methods from all the sources within the same plot (Figure 5). This agreement between the two methods has a slope of 1.30 [1.12, 1.49], $R = 0.94$, and $N = 42$ for emissions exceeding 500 kg/hr. The t-test p-value of the two methods is greater than 0.05 and suggests that the estimates from these two
400   methods came from the same distribution. Since the two methods agree for the blinded-volume controlled releases and the ~~flares~~ flare sources when the estimated emission rates are greater than 500 kg/hr, we infer that both methods ~~work~~ provide valid

emission estimates when emission rates are greater than 500 kg/hr. For rates below 500 kg/hr, both methods detect the sources down to 250 kg/hr, but the DI overestimates the rates due to the influence of nearby sources.

The mean percentage error of the methods should also be similar to those from the blinded-volume controlled release experiment (around 30%), given that uncertainty from both methods came largely from the influence of eddy-scale variability. The mIME simulation captures the variation over time in the simulations. The DI averages the flux information across various distances, capturing spatial variability due to eddies.

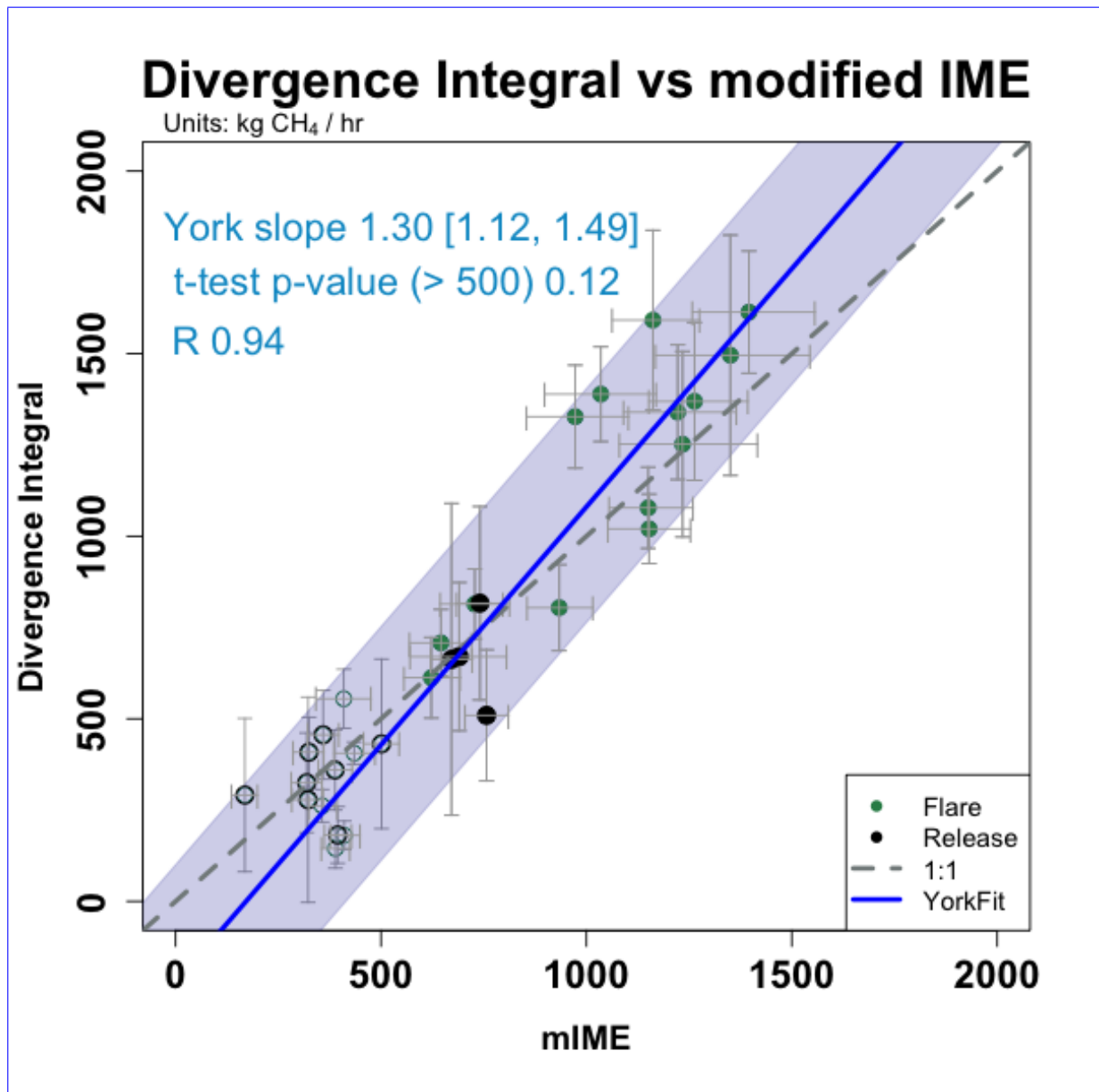



Figure 5. mIME vs DI estimated emissions for all valid cases from all the research flights of interest .in 2021. Using the algorithms presented here and meteorological fields from HRRR, MethaneAIR observations of XCH_4 enabled accurate estimation of methane emission from point sources within 30% errors, close to zero bias, and a detection limit of 200 kg/hr. The ~~red square represents no detection and flagged data below~~ black circles represent the ~~detection limit~~blind volume controlled release mIME estimates plotting against the DI emission estimates. The green circles represent the unlit flare mIME estimates plotting against the DI emission estimates. The circles representing emission estimates under 500 kg/hr are open. The circles representing emission estimates over 500 kg/hr are solid. The blue solid line is the post-unblinding York fit. The blue shaded area represents a 95% confidence interval of the York fit from resampling. The paired t-test between the ~~resamples~~Divergence Integral and the mIME emissions (p = 0.12, rate > 500 kg/hr) indicates no statistically significant difference between the population means for the two methods.

4   Conclusions

Our estimates of methane emissions from the blinded volume-controlled release experiment agreed well with the metered flow rates. The 2021 comparison between the MethaneAIR estimates and the known rates provides a slope of 0.96 [0.84, 1.08] ~~and suggests a~~ with an R of 0.83. The 2022 comparison offers a slope of 0.95 [0.81, 1.08] with an R of 0.94. The results suggest a 30% mean percentage error and ~~R of 0.83 and~~ a detection limit of 200 kg of methane per hr at 12960 m cruise altitude above ground (13850 m above sea level). The results validate the emission estimates for methane point sources from oil and gas infrastructure using XCH_4 data from the MethaneAIR imaging spectrometer, based on blinded-volume controlled release experiments and ~~a comparison of~~ comparing two algorithmic approaches for images of uncontrolled sources. Our study supports the application of our emission estimation algorithms to the upcoming MethaneSAT satellite after adjusting for the lower spatial resolution and higher signal-to-noise ratio, given appropriate testing and validation of the satellite. This study is limited by the number of collected data (21 data points on two days~~) from one locale~~, 43 points on four days) from two locales; further evaluations in a variety of conditions are needed to establish detection limits and accuracy more fully.

This study tackles the challenge of quantifying the uncertainty of emissions due to the effects of eddy-scale variability on remote sensing observations. Our methods explicitly estimate the influence of eddies on the values derived for each emission source of interest. The *mIME* captures the variation across simulated cycles. The DI averages the flux information from the XCH_4 image across spatial scales that capture plume size and shape modulation by eddy scale motions. From our limited data, our *mIME* and the original application of IME by Varon et al. (2018) perform equally well, provided that local wind information is available. Our *mIME* performs better than an IME using only mesoscale-scale winds from HRRR. Thus, the *mIME* provides an alternative to local wind data, which is typically not available, and it also produces error estimates associated with the wind variability near the source of interest.

We conclude that the new MethaneAIR imaging spectrometer supports accurate emission rate estimates for high emitting methane sources above 200 – 500 kg/hr detection limit when using the Integrated Mass Enhancement algorithm of Varon et al. (2018) modified for our system (*mIME*) or when applying our spatially oversampled application of the divergence integral. The sensor and analysis methods described here provide the greenhouse gas emission quantification community with improved approaches to consider the variability from the eddies. They establish the foundation for applying MethaneAIR and future MethaneSAT data to identify and quantify methane emissions to the atmosphere.

Code and data availability.   The datasets and code generated and analyzed during the current study are available at https://github.com/ju21u/mair_controlled_release.

Author contributions.   Author contributions. AC led this project. AC generated the LES, performed the mIME & ratio method, and wrote the manuscript with comments and revisions from all authors. SW supervised AC. MS led the DI analysis. TL helped AC develop the WRF-LES with additional LES insights from YL. SCW, MS, JSB, JEF, XL, KC, YL, RG, and MO helped to formulate the analyses. SCW, CCM, EC,

KC, AHS, JSW, BL, and SR prepared the MethaneAIR data used as inputs of quantification algorithms. RG and MO characterized facility-related bottom-up emission estimates. JSR, EDS, and AB were responsible for blinded-volume controlled release experiments on the ground. MS was responsible for the DI estimates. AC was responsible for the mIME estimates. AC prepared all figures for the manuscript. Everyone reviewed and edited the manuscript. Besides the first, second, and last, the authors were listed randomly and grouped by affiliations.

Competing interests. JSB, MO, and RG are employees of the Environmental Defense Fund (EDF). MethaneSAT LLC is a wholly-owned subsidiary of EDF. AC, MS, JEF, CCM, JSW, SR, and SWC received research funding from EDF

Acknowledgements. This work was supported by NSF EAGER grant 1856426 to Harvard University and by funding from the Environmental Defense Fund and Harvard University. The controlled methane release experiment was funded by ExxonMobil, the Stanford Strategic Energy Alliance, and the Stanford Natural Gas Initiative, an industry consortium that supports independent research at Stanford University. This work benefitted from discussions with S. Aminfard. We gratefully acknowledge the contributions of the flight crew and staff of NSF's Research Aviation Facility and discussions with the MethaneSAT Science Advisory Group and with S. Hamburg, Chief Scientist of the Environmental Defense Fund.

References

Bartholomew, J., Lyman, P., Weimer, C., and Tandy, W.: Wide area methane emissions mapping with airborne IPDA lidar, in: Lidar Remote Sensing for Environmental Monitoring 2017, vol. 10406, pp. 34–47, SPIE, https://doi.org/10.1117/12.2276713, 2017.

Blaylock, B. K., Horel, J. D., and Crosman, E. T.: Impact of Lake Breezes on Summer Ozone Concentrations in the Salt Lake Valley, Journal of Applied Meteorology and Climatology, 56, 353–370, https://doi.org/10.1175/JAMC-D-16-0216.1, 2017.

Chan Miller, C., Wofsy, Liu, and Roche: Methane retrieval from MethaneAIR using the CO2 Proxy Approach: A demonstration for the future MethaneSAT mission, 2022.

Conley, S., Franco, G., Faloona, I., Blake, D. R., Peischl, J., and Ryerson, T. B.: Methane emissions from the 2015 Aliso Canyon blowout in Los Angeles, CA, Science, 351, 1317–1320, https://doi.org/10.1126/science.aaf2348, publisher: American Association for the Advancement of Science, 2016.

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, https://eartharxiv.org/repository/view/3399/, publisher: EarthArXiv, 2022.

Conway, E. K., Souri, A. H., Benmergui, J., Sun, K., Liu, X., Staebell, C., Chan Miller, C., Franklin, J., Samra, J., Wilzewski, J., Roche, S., Luo, B., Chulakadabba, A., Sargent, M., Hohl, J., Daube, B., Gordon, I., Chance, K., and Wofsy, S.: Level0-to-Level1B processor for MethaneAIR, preprint, Gases/Remote Sensing/Data Processing and Information Retrieval, https://doi.org/10.5194/amt-2023-111, 2023.

Ehret, G., Bousquet, P., Pierangelo, C., Alpers, M., Millet, B., Abshire, J. B., Bovensmann, H., Burrows, J. P., Chevallier, F., Ciais, P., Crevoisier, C., Fix, A., Flamant, P., Frankenberg, C., Gibert, F., Heim, B., Heimann, M., Houweling, S., Hubberten, H. W., Jöckel, P., Law, K., 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, number: 10 Publisher: Multidisciplinary Digital Publishing Institute, 2017.

El Abbadi, S. H., Chen, Z., Burdeau, P. M., Rutherford, J. S., Chen, Y., Zhang, Z., Sherwin, E. D., and Brandt, A. R.: Comprehensive evaluation of aircraft-based methane sensing for greenhouse gas mitigation, https://eartharxiv.org/repository/view/5569/, publisher: EarthArXiv, 2023.

Etminan, M., Myhre, G., Highwood, E. J., and Shine, K. P.: Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing, Geophysical Research Letters, 43, https://doi.org/10.1002/2016GL071930, 2016.

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, Proceedings of the National Academy of Sciences, 113, 9734–9739, https://doi.org/10.1073/pnas.1605617113, iSBN: 9781605617114 Publisher: National Academy of Sciences Section: Physical Sciences, 2016.

Gaudet, B. J., Lauvaux, T., Deng, A., and Davis, K. J.: Exploration of the impact of nearby sources on urban atmospheric inversions using large eddy simulation, Elementa: Science of the Anthropocene, 5, 60, https://doi.org/10.1525/elementa.247, 2017.

Guanter, L., Irakulis-Loitxate, I., Gorroño, J., Sánchez-García, E., Cusworth, D. H., Varon, D. J., Cogliati, S., and Colombo, R.: Mapping methane point emissions with the PRISMA spaceborne imaging spectrometer, Remote Sensing of Environment, 265, 112 671, https://doi.org/10.1016/j.rse.2021.112671, 2021.

Irakulis-Loitxate, I., Guanter, L., Liu, Y.-N., Varon, D. J., Maasakkers, J. D., Zhang, Y., Chulakadabba, A., Wofsy, S. C., Thorpe, A. K., Duren, R. M., Frankenberg, C., Lyon, D. R., Hmiel, B., Cusworth, D. H., Zhang, Y., Segl, K., Gorroño, J., Sánchez-García, E., Sulprizio,

M. P., Cao, K., Zhu, H., Liang, J., Li, X., Aben, I., and Jacob, D. J.: Satellite-based survey of extreme methane emissions in the Permian basin, Science Advances, 7, eabf4507, https://doi.org/10.1126/sciadv.abf4507, 2021.

490 Jacob, D. J., Turner, A. J., Maasakkers, J. D., Sheng, J., Sun, K., Liu, X., Chance, K., Aben, I., McKeever, J., and Frankenberg, C.: Satellite observations of atmospheric methane and their value for quantifying methane emissions, Atmospheric Chemistry and Physics, 16, 14 371–14 396, https://doi.org/10.5194/acp-16-14371-2016, 2016.

Jervis, D., McKeever, J., Durak, B. O. A., Sloan, J. J., Gains, D., Varon, D. J., Ramier, A., Strupler, M., and Tarrant, E.: The GHGSat-D imaging spectrometer, Atmospheric Measurement Techniques, 14, 2127–2140, https://doi.org/10.5194/amt-14-2127-2021, publisher:

495 Copernicus GmbH, 2021.

Johnson, M. R., Tyner, D. R., and Szekeres, A. J.: Blinded evaluation of airborne methane source detection using Bridger Photonics LiDAR, Remote Sensing of Environment, 259, 112 418, https://doi.org/10.1016/j.rse.2021.112418, 2021.

Jongaramrungruang, S., Frankenberg, C., Matheou, G., Thorpe, A. K., Thompson, D. R., Kuai, L., and Duren, R. M.: Towards accurate methane point-source quantification from high-resolution 2-D plume imagery, Atmospheric Measurement Techniques, 12, 6667–6681,

500 https://doi.org/10.5194/amt-12-6667-2019, 2019.

Laboratory, U.-E. O.: NSF/NCAR GV HIAPER Aircraft, https://doi.org/10.5065/D6DR2SJP, publisher: UCAR/NCAR - Earth Observing Laboratory, 2005.

Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., Sweeney, C., Diaz, L. I., Martins, D., Shepson, P. B., and Davis, K. J.: Constraining the CO_2 budget of the corn belt: exploring uncertainties from the assumptions in a

505 mesoscale inverse system, Atmospheric Chemistry and Physics, 12, 337–354, https://doi.org/10.5194/acp-12-337-2012, 2012.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Zhang, H., Aamaas, B., Boucher, O., Dalsøren, S. B., Daniel, J. S., Forster, P., Granier, C., Haigh, J., Hodnebrog, , Kaplan, J. O., Marston, G., Nielsen, C. J., O'Neill, B. C., Peters, G. P., Pongratz, J., Ramaswamy, V., Roth, R., Rotstayn, L., Smith, S. J., Stevenson, D., Vernier, J.-P., Wild, O., Young, P., Jacob, D., Ravishankara, A. R., and Shine, K.: 8 Anthropogenic and Natural Radiative

510 Forcing, p. 82, 2013.

Nakanishi, M. and Niino, H.: An Improved Mellor–Yamada Level-3 Model: Its Numerical Stability and Application to a Regional Prediction of Advection Fog, Boundary-Layer Meteorology, 119, 397–407, https://doi.org/10.1007/s10546-005-9030-8, 2006.

Nakanishi, M. and Niino, H.: Development of an Improved Turbulence Closure Model for the Atmospheric Boundary Layer, Journal of the Meteorological Society of Japan. Ser. II, 87, 895–912, https://doi.org/10.2151/jmsj.87.895, 2009.

515 OCD: Notification Of Release, 372422 (44680), https://ocdimage.emnrd.nm.gov/imaging/IncidentFileView.aspx?incident=nAPP2123850791, 2021.

Rutherford, J. S., Sherwin, E. D., Chen, Y., Aminfard, S., and Brandt, A. R.: Evaluating methane emission quantification performance and uncertainty of aerial technologies via high-volume single-blind controlled releases, https://eartharxiv.org/repository/view/5113/, publisher: EarthArXiv, 2023.

520 Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., Raymond, P. A., Dlugokencky, E. J., Houweling, S., Patra, P. K., Ciais, P., Arora, V. K., Bastviken, D., Bergamaschi, P., Blake, D. R., Brailsford, G., Bruhwiler, L., Carlson, K. M., Carrol, M., Castaldi, S., Chandra, N., Crevoisier, C., Crill, P. M., Covey, K., Curry, C. L., Etiope, G., Frankenberg, C., Gedney, N., Hegglin, M. I., Höglund-Isaksson, L., Hugelius, G., Ishizawa, M., Ito, A., Janssens-Maenhout, G., Jensen, K. M., Joos, F., Kleinen, T., Krummel, P. B., Langenfelds, R. L., Laruelle, G. G., Liu, L., Machida, T., Maksyutov, S., McDonald, K. C., McNorton, J., Miller, P. A., Melton,

525 J. R., Morino, I., Müller, J., Murguia-Flores, F., Naik, V., Niwa, Y., Noce, S., O'Doherty, S., Parker, R. J., Peng, C., Peng, S., Peters, G. P.,

Prigent, C., Prinn, R., Ramonet, M., Regnier, P., Riley, W. J., Rosentreter, J. A., Segers, A., Simpson, I. J., Shi, H., Smith, S. J., Steele, L. P., Thornton, B. F., Tian, H., Tohjima, Y., Tubiello, F. N., Tsuruta, A., Viovy, N., Voulgarakis, A., Weber, T. S., van Weele, M., van der Werf, G. R., Weiss, R. F., Worthy, D., Wunch, D., Yin, Y., Yoshida, Y., Zhang, W., Zhang, Z., Zhao, Y., Zheng, B., Zhu, Q., Zhu, Q., and Zhuang, Q.: The Global Methane Budget 2000–2017, Earth System Science Data, 12, 1561–1623, https://doi.org/10.5194/essd-12-1561-2020, publisher: Copernicus GmbH, 2020.

Schneising, O., Buchwitz, M., Reuter, M., Vanselow, S., Bovensmann, H., and Burrows, J. P.: Remote sensing of methane leakage from natural gas and petroleum systems revisited, Atmospheric Chemistry and Physics, 20, 9169–9182, https://doi.org/10.5194/acp-20-9169-2020, 2020.

Sherwin, E. D., Chen, Y., Ravikumar, A. P., and Brandt, A. R.: Single-blind test of airplane-based hyperspectral methane detection via controlled releases, Elementa: Science of the Anthropocene, 9, 00 063, https://doi.org/10.1525/elementa.2021.00063, 2021.

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 methane emissions detection and quantification, https://eartharxiv.org/repository/view/3465/, publisher: EarthArXiv, 2022.

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, Scientific Reports, 13, 3836, https://doi.org/10.1038/s41598-023-30761-2, number: 1 Publisher: Nature Publishing Group, 2023.

Shindell, D., Kuylenstierna, J. C. I., Vignati, E., van Dingenen, R., Amann, M., Klimont, Z., Anenberg, S. C., Muller, N., Janssens-Maenhout, G., Raes, F., Schwartz, J., Faluvegi, G., Pozzoli, L., Kupiainen, K., Höglund-Isaksson, L., Emberson, L., Streets, D., Ramanathan, V., Hicks, K., Oanh, N. T. K., Milly, G., Williams, M., Demkine, V., and Fowler, D.: Simultaneously Mitigating Near-Term Climate Change and Improving Human Health and Food Security, Science, 335, 183–189, https://doi.org/10.1126/science.1210026, publisher: American Association for the Advancement of Science, 2012.

Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Barker, D., Duda, G., Huang, X.-y., Wang, W., and Powers, G.: A Description of the Advanced Research WRF Version 3, https://doi.org/10.5065/D68S4MVH, 2008.

Staebell, C., Sun, K., Samra, J., Franklin, J., Chan Miller, C., Liu, X., Conway, E., Chance, K., Milligan, S., and Wofsy, S.: Spectral calibration of the MethaneAIR instrument, Atmospheric Measurement Techniques, 14, 3737–3753, https://doi.org/10.5194/amt-14-3737-2021, publisher: Copernicus GmbH, 2021.

Sun, K., Zhu, L., Cady-Pereira, K., Chan Miller, C., Chance, K., Clarisse, L., Coheur, P.-F., González Abad, G., Huang, G., Liu, X., Van Damme, M., Yang, K., and Zondlo, M.: A physics-based approach to oversample multi-satellite, multispecies observations to a common grid, Atmospheric Measurement Techniques, 11, 6679–6701, https://doi.org/10.5194/amt-11-6679-2018, publisher: Copernicus GmbH, 2018.

Sánchez-García, E., Gorroño, J., Irakulis-Loitxate, I., Varon, D. J., and Guanter, L.: Mapping methane plumes at very high spatial resolution with the WorldView-3 satellite, preprint, Gases/Remote Sensing/Data Processing and Information Retrieval, https://doi.org/10.5194/amt-2021-238, 2021.

US Department of Commerce, N.: ASOS, https://www.weather.gov/asos/, publisher: NOAA's National Weather Service.

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, Atmospheric Measurement Techniques, 11, 5673–5686, https://doi.org/10.5194/amt-11-5673-2018, 2018.

Varon, D. J., Jervis, D., McKeever, J., Spence, I., Gains, D., and Jacob, D. J.: High-frequency monitoring of anomalous methane point sources with multispectral Sentinel-2 satellite observations, Atmospheric Measurement Techniques, 14, 2771–2785, https://doi.org/10.5194/amt-14-2771-2021, publisher: Copernicus GmbH, 2021.

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote Sensing of Environment, 120, 70–83, https://doi.org/10.1016/j.rse.2011.09.027, 2012.

Supplemental Materials:
Methane Point Source Quantification Using MethaneAIR: A New Airborne Imaging Spectrometer

August 23, 2023

S1   Calculations

S1.1   mIME Effective Wind Speeds

The effective wind speeds used in the mIME calculation came from the relationship proposed by Varon et al. (2018), $U_{eff} = \alpha \cdot log(10\,m\,wind) + 0.6$ where α is between 0.9-1.1. We took advantage of the LES runs by introducing the LES-specific effective wind $U_{adaptive,eff}$ for each time step. $U_{adaptive,eff}$ is defined as $\frac{Q \cdot L}{IME}$. Because we know all the terms in this equation from the LES, we can calculate the true $U_{adaptive,eff}$ that gives the best estimate of Q.

By calculating the $U_{adaptive,eff}$ for all the time steps of interest, we end up with multiple pairs of $U_{adaptive,eff}$ and U, which can be fitted into a linear regression function of the form $U_{eff} = a \cdot log(10\,m\,wind) + b$ with unique a and b coefficients.

~~However, the~~ The LES-specific relationship can introduce overfitting results when the data points are limited, and wind speeds are low. If we overfit the relationship, we cannot justify the predicted effective wind speeds and may come up with biased emission estimates. Since the wind speeds during the controlled release experiments were lower than 5 m/s for most of the time, we decided to use the coefficients in (Varon et al., 2018) ~~to avoid overfitting~~ for the sake of simplicity.

Though the coefficients (Varon et al., 2018) derived from idealized LES with varying meteorological conditions are the most comprehensive IME coefficients available, we should be aware of ambient weather and turbulence conditions that may affect the relationship's universality.

S1.2   Divergence Integral

The divergence integral method is based on the integral form of the continuity equation, which states that the net production of CH_4 inside an enclosing surface, $\int_V P$, through an enclosing surface is given by

$$\int_V P = \int_V \vec{\nabla} \cdot c\vec{v} + \int_V \frac{dm}{dt} = 0 \tag{S1}$$

where c is the ~~concentration of methane~~ methane concentration, v is the wind speed, and $\langle \frac{\partial m}{\partial t} \rangle$ is the change in mass of methane with time within volume V. $\int_V P = 0$ since there is no *in situ* production. The flux divergence can be decomposed into two terms:

$$\nabla \cdot cv = v \cdot \nabla c + c \cdot \nabla v \tag{S2}$$

Conley et al. (2017) demonstrate that for wind speeds over 1 m/s and fluxes above 200 kg/hr, the wind divergence term (2nd term in Eq. S2) near a source is 1% or less of the concentration gradient term (1st term in Eq. S2); therefore, we can neglect the 2nd term for our application. To isolate the XCH$_4$ enhancement relative to the surroundings, we subtract the mean column XCH$_4$ along the rectangle from the divergence term, which does not alter ∇c, because $\nabla \langle c \rangle = 0$

$$c' = c - \langle c \rangle \tag{S3}$$

$$\nabla c = \nabla(\langle c \rangle + c') = \nabla c' \tag{S4}$$

where c' is the XCH$_4$ enhancement and $\langle c \rangle$ is the mean XCH$_4$ around the rectangle. ~~For MethaneAIR data,~~ The c' ~~was calculated from the measured XCH$_4$ as follows:~~

$$c' = (XCH_4 - \langle XCH_4 \rangle_{rect.}) * n_{column} * M_{CH4}$$

~~where n_{column} is the moles of air in the column based on the surface pressure from HRRR and M_{CH4} is the molar mass of methane~~values in S3 equation can be both positive and negative. When there is a positive flux through the surface, the positive c' values outweigh the negative c' values. There are some instances where the flux through the surface is negative (i.e., Figure 2d, 420m), and those values are averaged into the total flux as well. In most cases, these reflect the influence of boundary layer eddies.

For analysis of the total column data from MethaneAIR in the vicinity of an emission source, we assume that methane concentrations above the planetary boundary layer are equal to unperturbed values outside the bounding surface. Then, the horizontal gradients of XCH_4 arise entirely in the boundary layer, and we may take v ~~to be~~ as the pressure-weighted mean boundary layer wind speed and direction.

We then use Gauss's theorem to relate the volume integral to a surface integral around the cuboid:

$$
\begin{aligned}
0 &= \int_V v \cdot \nabla c' dV + \langle \frac{\partial m}{\partial t} \rangle \\
&= -\Phi_{surf} + \oint_{\partial V = S} v \cdot \hat{n} c' dS + \langle \frac{\partial m}{\partial t} \rangle \\
&= -\Phi_{surf} + \oint \partial V = S v \cdot \hat{n}((XCH_4 - \langle XCH_4 \rangle_{rect.}) * n_{column} * M_{CH4}) dS + \langle \frac{\partial m}{\partial t} \rangle
\end{aligned}
$$
(S5)

$$\Phi_{surf} = \oint \partial V = S v \cdot \hat{n}((XCH_4 - \langle XCH_4 \rangle_{rect.}) * n_{column} * M_{CH4}) dS$$

where Φ_{surf} is the flux into the volume from the land surface, S is the surface enclosing the cuboid~~and~~, $\hat{n}$ is an outward pointing unit vector normal to the surface, n_{column} is the moles of air in the column based on the surface pressure from HRRR and M_{CH4} is the molar mass of methane. For a volume the size of a single plume, the term $\langle \frac{\partial m}{\partial t} \rangle$ should be insignificant, assuming that the plume is in a steady state and no mass is building up in the volume. We assume that the vertical flux through the top of the cuboid at the planetary boundary layer (PBL) height is zero, consistent with neglecting the divergence of the horizontal wind. Thus, ~~we assume that~~ the horizontal flux through the surface is balanced by and therefore equal to the flux through the bottom of the cuboid from the surface source, Φ_{surf} ~~,~~(kg/hr), we wish to measure. ~~Combining equations (A5) and (A6) and~~Based on Eq. S5, the flux from the source is computed from observed XCH_4 as follows:

$$\Phi_{surf} = \sum^{\text{around rect}} (XCH_{4_i} - \langle XCH_4 \rangle_{rect.}) \cdot n_{\text{column}} \cdot M_{CH_4} \cdot v \cdot \hat{n} \cdot \Delta l$$
(S6)

where Δl is the distance between successive XCH_4 measurements. We average the fluxes calculated over a sequence of rectangles ~~,~~extending downwind of the source, spanning several eddy scales. The box grows by 1 pixel in each step except for the upwind direction, where it moves over by 1/4 pixel. The upwind boundary was selected to be close to the upwind of the source using the moment of inertia of the plume (see Fig. 2)

S1.3 The Ratio Method

We developed a new method based on WRF-LES-HRRR outputs, assuming that the LES plumes represented the actual plumes in both areas and concentration distributions within the plume. At a given location and time, the probabilistic distribution of methane plumes should be similar at any emission rate. The emission rates should linearly depend on the magnitude of concentrations, which can be written as

$$\frac{Q_{unknown}}{Q_{known}} = \frac{IME_{unknown}}{IME_{known}}.$$

IME_{know} refers to the values from the LES simulations with a nominal emission rate (usually 988 kg/hr). Thus, the emission rates of the observed plumes are a function of the nominal emission rate of the simulated plumes multiplied by the ratio of the total masses of the observed plumes to the simulated plumes. The relationship can be expressed as

$$Q_{unknown} = Q_{known} \cdot \frac{IME_{unknown}}{IME_{known}}.$$
(S7)

This ratio method was first used in (Irakulis-Loitxate et al., 2021). It is applicable to very large sources with plumes large enough to be well represented by the 111 m × 111 m resolution. This is the limitation of the method that it only works when simulated winds accurately represent the actual winds.

Since LES cannot capture the shape of a snapshot of the observed plume due to large stochastic variations of the atmosphere, we use a range of snaps across time to represent the observed plumes for these two LES methods. We took the standard deviation of the emission rates calculated from 7 snapshots across a 30 minutes interval (15 minutes before and 15 minutes after the retrieval), 5 minutes apart from one another, to avoid autocorrelation effects. For the MethaneAIR data from RF04 and RF05 flights, only a few plumes were large enough to be assessed with the ratio and this method is not used in our post-unblinding analysis.

S1.4   Gridding by Nearest Neighbor and Physics-Based Methods

The nearest neighbor method for gridding used the R function binMean from the oce package to place the georeferenced XCH_4 data onto a predefined 10m $\times$ 10m grid. The retrieved XCH_4 values have a standard deviation of about 80 ppb. Since the the point spread function is about 12 m wide across the track, and 35 m along track, the grid of native pixels spatially oversamples the image on the ground. Therefore, we spatially filtered the retrieved XCH_4 data to create the gridded product in two steps. First, the holes in the initial gridded field were filled using a 3x3 averaging matrix. Then we further spatially filtered this field by a Gaussian filter with full width at half maximum of 1.7 pixels before inputting the XCH_4 values to our point source quantification methods. The resulting oversampled gridded product has a standard deviation of 40 ppb and a spatial correlation length of about 20 m, comparable to the length scale of the point spread function. The nearest-neighbor gridding method was used for the initial (blinded) input from Harvard to Stanford.

The improved estimates after unblinding used the "physics-based" gridding method of Sun et al. (2018).

S1.5   York Regression

When errors in the observables are present on both the x-axis and the y-axis, it is recommended to use York regression and other equations derived from York to determine the slope, y-intercept, and errors of the least-square estimation solutions. The full equations and derivations can be found on York (1968) and York et al. (2004)

S1.6   Mean Percentage Error

$$MPE = \frac{100\%}{n} \sum_{t=1}^{n} \frac{a_t - f_t}{a_t} \tag{S8}$$

where a_t is the actual value of the quantity being forecast, f_t is the forecast, and n is the number of different times for which the variable is forecast.

S2   Additional Results

S2.1   Blinded Volume-Controlled Release Experiments

The release rate was measured by a Sierra 640i QuadraTherm meter. Stanford performed an extensive set of meter intercomparison tests to characterize uncertainty in the QuadraTherm 640i meters (Rutherford et al. (2023)). Based on these intercomparison tests, a bias of up to 6% can be observed between QuadraTherm 640i meters. Error in metered controlled release volumes is characterized using a Monte Carlo analysis that also accounts for uncertainty in gas composition.

On seven occasions during the missions, the Stanford team informed the MethaneAIR team of the wind direction at the site since winds were light and variable, as reported by nearby ASOS systems. The MethaneAIR team considered reorienting the flight track to capture the plumes due to the wind conditions, but no changes in the flight track were required. On two occasions, Stanford confirmed that local winds were not zero. Although null winds were not desired for this first test of MethaneAIR, since the winds were not zero, the decision was made to complete the surveys as planned.

For the analysis, the MethaneAIR team ~~preferred to use~~ used the York fit, which accounts for uncertainties in both the X and the Y coordinates, instead of the commonly used OLS, which ~~only accounts for uncertainties~~ assumes that uncertainties are significant only in the Y coordinate. ~~York fit allows the remaining errors (besides mIME quantification errors and the metered errors) to be weighted in both coordinates. Thus~~In this experiment, the uncertainties in the metered flow rates are not negligible, and the error estimates are available from the Stanford team. Hence, the York fit ~~should be, by design, closer to~~ was selected to approximate the desired Maximum Likelihood Estimator ~~value~~ for the relationship between the Harvard values and the Stanford emission rates.

The MethaneAIR team conducted tests on the OLS to verify whether the OLS is an appropriate regression choice. The slope of the OLS line relating the Harvard mIME emissions to the Stanford blinded-volume controlled release rates is 0.78 for the original submission. Based on the heteroskedasticity test, this OLS fit violates the assumptions for linear regression (p-value 0.006). The violation ~~suggests~~ indicates a significant relationship between the residuals and the predictor (Stanford emission rates). ~~However, when using the~~ The Non-Constant Error Variance test (ncvTest) ~~, the~~ gave a p-value ~~is~~ of 0.14, suggesting ~~significant violation only~~ violation at the 84% level. These tests support the application of the York fit for the 2021 results.

Based on the 2021 experiment (N = 21), the false positive rate is $\approx 10\%$ and the false negative is $\approx 5\%$. These rates can not be generalized to the instrument capability since we have a limited number of data points, and the false positive/negative rates depend on the scene of interest.

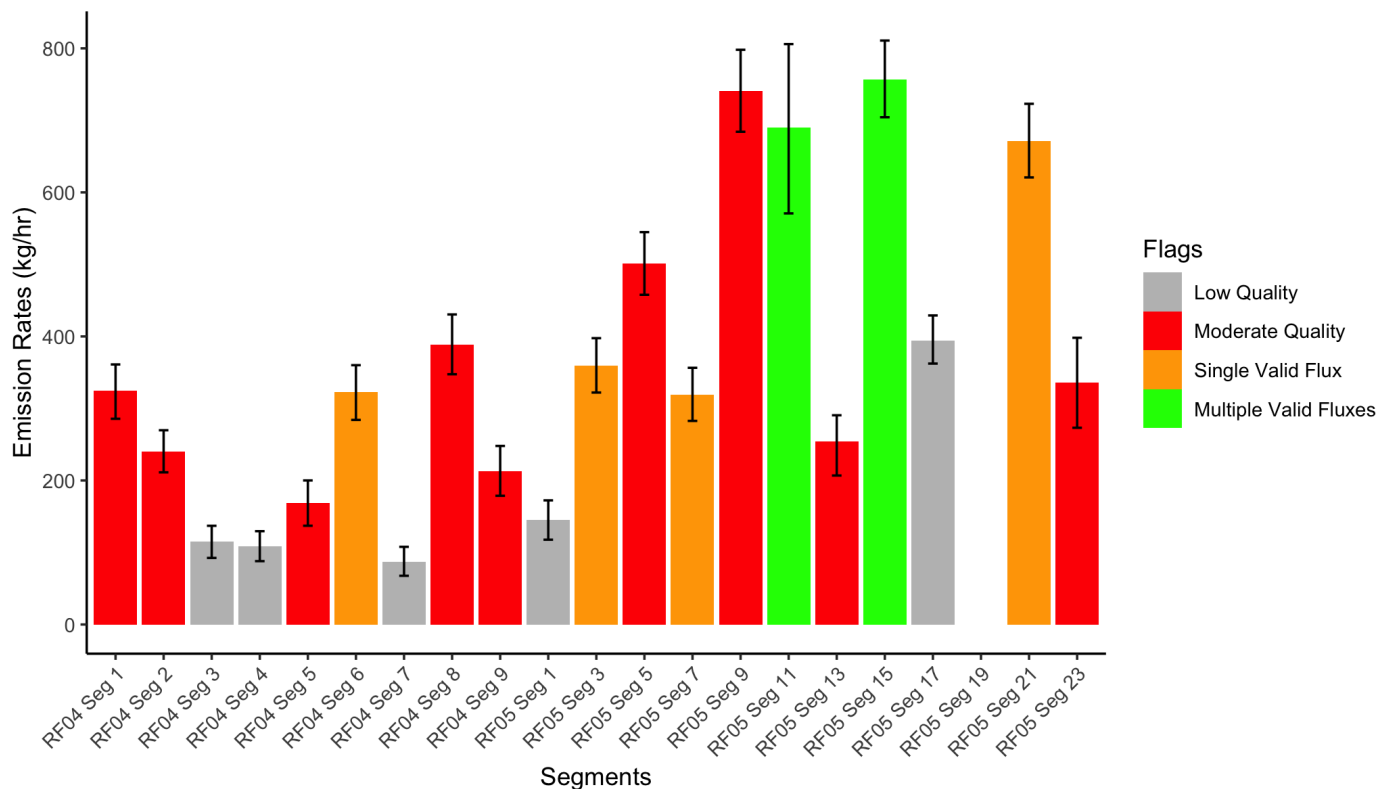



Figure S1: Post-unblinding best estimates of methane emissions from the controlled release experiments during RF04 and RF05 using the mIME method. The endpoints are tagged and color-coded according to the Decision Tree in section S10.

Table S1: RF04 and RF05 combined best estimate results from the blinded-volume controlled release experiments (32.053 °N, 102.301 °W) and MethaneAIR Level 3 Gaussian filtered data on 30 July 2021 (RF04) and 3 August 2021 (RF05). The times are in UTC. The emission rates, lower bounds, and upper bounds are in kg/hr. The blinded, blinded lower, and blinded upper columns were reported to the Stanford team as MethaneAIR initial emission estimates on 1 February 2022. The unblinded, unblinded lower, and unblinded upper columns were added after the unblinding process. The upper and lower bound estimates were calculated using the method described in Section 2.5.

Seg	Timestamp (UTC)	Blinded	Blinded Lower	Blinded Upper	Unblinded	Unblinded Lower	Unblinded Upper	Flagged
1	30/7/21 15:41	321.74	179.30	471.71	324.55	285.64	361.07	No
2	30/7/21 15:52	232.19	133.63	338.38	240.14	211.28	269.73	No
3	30/7/21 16:07	109.22	62.73	160.03	114.76	92.43	137.06	Yes
4	30/7/21 16:22	102.76	59.80	151.54	108.16	87.96	129.54	Yes
5	30/7/21 16:36	152.31	90.07	222.33	168.53	137.12	200.03	No
6	30/7/21 16:51	225.42	99.69	392.61	322.45	284.06	360.08	No
7	30/7/21 17:06	82.21	33.30	136.95	87.43	67.63	107.83	Yes
8	30/7/21 17:22	362.67	138.79	613.80	388.06	347.49	430.51	No
9	30/7/21 18:08	205.48	89.54	331.16	213.25	178.73	247.93	Yes
1	3/8/21 15:53	140.85	76.46	211.86	140.85	76.46	211.86	Yes
3	3/8/21 16:11	344.58	189.55	512.42	344.58	189.55	512.42	No
5	3/8/21 16:30	472.22	280.28	682.46	472.22	280.28	682.46	No
7	3/8/21 16:49	299.01	197.33	415.01	299.01	197.33	415.01	No
9	3/8/21 17:06	694.32	537.16	859.69	694.32	537.16	859.69	No
11	3/8/21 17:24	648.93	521.87	791.07	648.93	521.87	791.07	No
13	3/8/21 17:41	243.47	166.76	326.58	243.47	166.76	326.58	Yes
15	3/8/21 17:59	730.75	494.07	987.37	730.75	494.07	987.37	No
17	3/8/21 18:15	385.89	288.5	487.62	385.89	288.50	487.62	Yes
19	3/8/21 18:35	NA	NA	NA	NA	NA	NA	No
21	3/8/21 18:53	525.67	342.93	727.14	525.67	342.93	727.14	No
23	3/8/21 19:11	173.21	118.33	231.18	173.21	118.33	231.18	No

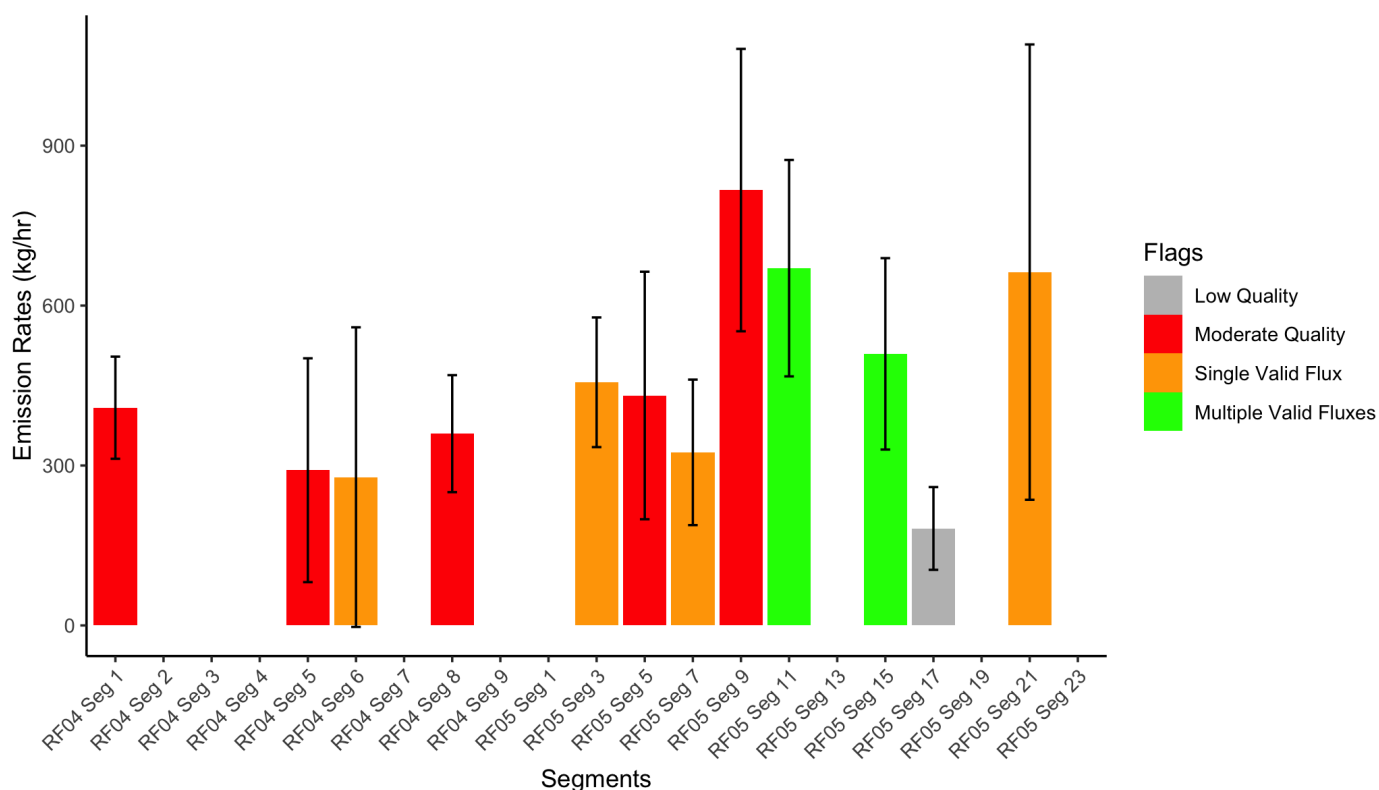



Figure S2: Post-unblinding methane emission estimates from the blinded-volume controlled release experiments during RF04 and RF05 using the DI method. The endpoints are tagged and color-coded according to the Decision Tree in section S10.

Table S2: DI results from the blinded-volume controlled release experiments (32.053 °N, 102.301 °W) and MethaneAIR Level 3 Gaussian filtered data on 30 July 2021 and 3 August 2021. The times are in UTC. The emission rates, lower bounds, and upper bounds are in kg/hr.

Seg	Timestamp (UTC)	Unblinded Estimate	Lower Bound	Upper Bound
1	30/7/21 15:41	408.43	312.61	504.25
2	30/7/21 15:52	NA	NA	NA
3	30/7/21 16:07	NA	NA	NA
4	30/7/21 16:22	NA	NA	NA
5	30/7/21 16:36	291.14	81.20	501.09
6	30/7/21 16:51	278.21	-2.83	559.24
7	30/7/21 17:06	NA	NA	NA
8	30/7/21 17:22	359.74	250.00	469.49
9	30/7/21 18:08	NA	NA	NA
1	3/8/21 15:53	NA	NA	NA
3	3/8/21 16:11	456.12	334.50	577.73
5	3/8/21 16:30	431.36	199.19	663.53
7	3/8/21 16:49	324.64	188.14	461.15
9	3/8/21 17:06	816.66	551.81	1081.50
11	3/8/21 17:24	670.18	467.17	873.18
13	3/8/21 17:41	NA	NA	NA
15	3/8/21 17:59	509.45	329.93	688.98
17	3/8/21 18:15	181.93	104.31	259.54
19	3/8/21 18:35	NA	NA	NA
21	3/8/21 18:53	662.77	235.70	1089.84
23	3/8/21 19:11	NA	NA	NA

Table S3: Original results for all methods from the blinded-volume controlled release experiments (32.053 °N, 102.301 °W) and MethaneAIR Level 3 Gaussian filtered data on 30 July 2021 and 3 August 2021. The emission rates, lower bounds, and upper bounds are in kg/hr. This table was reported to the Stanford team as MethaneAIR initial emission estimates on 1 February 2022. The times are in UTC. The bounds were calculated using the method described in Section 2.5. The SNR is the signal-to-noise ratio. The SL is the scale length of each plume in meters

Seg	Timestamp (UTC)	IME	IME.l	IME.u	Ratio	Ratio.l	Ratio.u	Gaussian	Gauss.l	Gauss.u	SNR	SL
1	30/7/21 15:41	321.74	312.25	330.84	1267.08	1072.46	1419.36	387.04	318.53	455.55	2.33	153.92
2	30/7/21 15:52	232.19	222.98	241.58	765.04	631.16	885.97	NA	NA	4.44	2.45	128.82
3	30/7/21 16:07	109.22	102.16	115.46	206.19	159.37	250.17	NA	NA	NA	2.19	79.98
4	30/7/21 16:22	102.76	94.56	110.81	224.24	189.35	255.61	NA	NA	NA	2.18	87.16
5	30/7/21 16:36	152.31	137.82	165.12	362.74	307	412.7	253.13	158.86	347.39	2.26	106.28
6	30/7/21 16:51	291.55	266.56	313.72	572.28	448.29	685.25	159.29	83.16	235.42	2.5	175.46
7	30/7/21 17:06	82.21	75.89	87.72	131.26	109.96	144.92	NA	NA	NA	2.43	58.3
8	30/7/21 17:22	362.67	339.8	385.66	694.59	589.45	812.25	296.98	260.15	333.82	2.45	196.69
9	30/7/21 18:08	205.48	196.33	214.49	515.09	316.43	651.97	98.32	14.51	182.13	2.13	115.74
1	3/8/21 15:53	140.85	133.45	148.46	215.43	190.36	240.07	NA	NA	NA	2.25	85.46
3	3/8/21 16:11	344.58	322.5	362.84	819.55	664.12	965.76	282.12	223.41	340.84	2.53	185.53
5	3/8/21 16:30	472.22	445.41	497.88	806.07	690.37	972.87	258.84	204.09	313.59	2.49	237.98
7	3/8/21 16:49	299.01	275.76	323.11	707.65	589.56	799.42	128.58	89.51	167.65	2.62	185.79
9	3/8/21 17:06	694.32	644.67	736.81	752.41	616.16	870.63	784.38	710.68	858.07	2.29	363.56
11	3/8/21 17:24	648.93	587.62	711.46	1130.2	784.58	1471.68	1056.21	981.26	1131.17	2.75	203.28
13	3/8/21 17:41	243.47	229.22	256.64	932.64	759.34	1083.86	313.8	252.26	375.35	2.02	128.49
15	3/8/21 17:59	730.75	701.77	761.88	1189.06	1078.02	1297.29	192.44	90.55	294.34	2.52	280.44
17	3/8/21 18:15	385.89	375.8	395.69	818.72	687.49	918.82	NA	NA	3.18	2.48	199.81
19	3/8/21 18:35	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
21	3/8/21 18:53	525.67	490.81	558.78	1242.4	1020.6	1441.07	328.44	207.52	449.37	2.83	171.8
23	3/8/21 19:11	173.21	167.15	178.92	210	178.99	234.64	505.73	447.7	563.76	2.39	53.87

Table S4: Blinded-volume controlled release data reported by the Stanford team on 1 February 2022. The cr_kgh_CH4_mean90, cr_kgh_CH4_lower90, and cr_kgh_CH4_upper90 represent the 90-second window rolling average, lower, and upper flow rates in kg/hr. The wind speeds are in m/s. The timestamps are in UTC.

Seg	Timestamp (UTC)	cr_kgh_CH4_mean90	cr_kgh_CH4_lower90	cr_kgh_CH4_upper90	Winds
1	30/7/21 15:41	358.48	334.29	384.75	2.8
2	30/7/21 15:52	171.08	154.32	188.05	2.71
3	30/7/21 16:07	49.98	36.06	63.71	2.44
4	30/7/21 16:22	0.00	0.00	0.00	2.29
5	30/7/21 16:36	236.74	217.42	256.81	2.36
6	30/7/21 16:51	295.08	273.58	318.53	2.59
7	30/7/21 17:06	95.05	80.53	109.95	2.68
8	30/7/21 17:22	382.25	356.59	409.77	2.54
9	30/7/21 18:08	63.65	53.99	73.67	3.29
1	3/8/21 15:52	6.72	-2.07	15.42	2.32
3	3/8/21 16:11	495.05	464.61	528.52	2.28
5	3/8/21 16:30	560.57	527.17	596.61	2.27
7	3/8/21 16:48	297.28	277.02	318.78	2.03
9	3/8/21 17:06	654.38	615.38	695.86	1.92
11	3/8/21 17:24	690.80	648.48	734.36	2.07
13	3/8/21 17:41	0.00	0.00	0.00	2
15	3/8/21 17:59	382.28	356.57	409.38	1.64
17	3/8/21 18:15	129.63	113.98	145.39	1.87
19	3/8/21 18:35	56.05	41.93	70.22	2.48
21	3/8/21 18:53	446.05	417.23	476.86	2.42
23	3/8/21 19:11	236.63	217.37	256.66	2.21

Table S5: Blinded-volume controlled release data reported by the Stanford team on 22 March 2023. The release_rate_kgh is the metered rate reported by the Stanford team. The lower_95CI and upper_95CI are the lower and upper bound of the 95th percentile confidence interval of the metered rates. The MAIR is our blinded best estimate of emissions reported to the Stanford team. The MAIR_lower and MAIR_upper are the lower and upper bound of the 95th percentile confidence interval of our estimates. The timestamps are in UTC.

Seg	Timestamp (UTC)	release_rate_kgh	lower_95CI	upper_95CI	MAIR	MAIR_lower	MAIR_upper
1	25/10/22 17:15	205.23	203.92	206.55	177	121	232
2	25/10/22 17:32	97.04	96.42	97.66	61	18	105
3	25/10/22 17:50	987.13	936.13	1038.13	855	578	1133
4	25/10/22 18:09	982.50	928.47	1036.52	908	662	1155
5	25/10/22 18:27	0.00	0.00	0.00	0	-100	100
6	25/10/22 18:47	468.33	463.82	472.85	669	409	928
7	25/10/22 19:07	24.42	24.31	24.53	0	-100	100
8	25/10/22 19:28	634.57	625.46	643.68	544	374	714
9	25/10/22 20:10	142.14	141.52	142.75	303	183	423
10	25/10/22 20:29	74.65	74.32	74.99	89	-11	189
11	25/10/22 20:46	63.03	62.78	63.29	140	89	191
1	29/10/22 16:25	0.00	0.00	0.00	0	-100	100
2	29/10/22 16:50	340.16	336.53	343.79	196	97	294
3	29/10/22 17:17	671.74	660.33	683.16	787	290	1283
4	29/10/22 17:42	405.59	400.31	410.87	644	271	1017
5	29/10/22 18:07	33.61	33.27	33.94	94	-6	194
6	29/10/22 18:31	337.56	333.89	341.23	388	296	480
7	29/10/22 18:54	818.30	793.86	842.75	958	549	1368
8	29/10/22 19:16	1060.77	1002.64	1118.91	1199	886	1512
9	29/10/22 19:38	1289.35	1221.55	1357.15	1489	1004	1974
10	29/10/22 19:59	585.32	575.53	595.11	579	397	761
11	29/10/22 20:21	199.26	165.06	233.47	252	147	357
12	29/10/22 20:42	493.52	487.02	500.02	584	328	839
13	29/10/22 21:02	168.86	167.19	170.53	112	-21	244

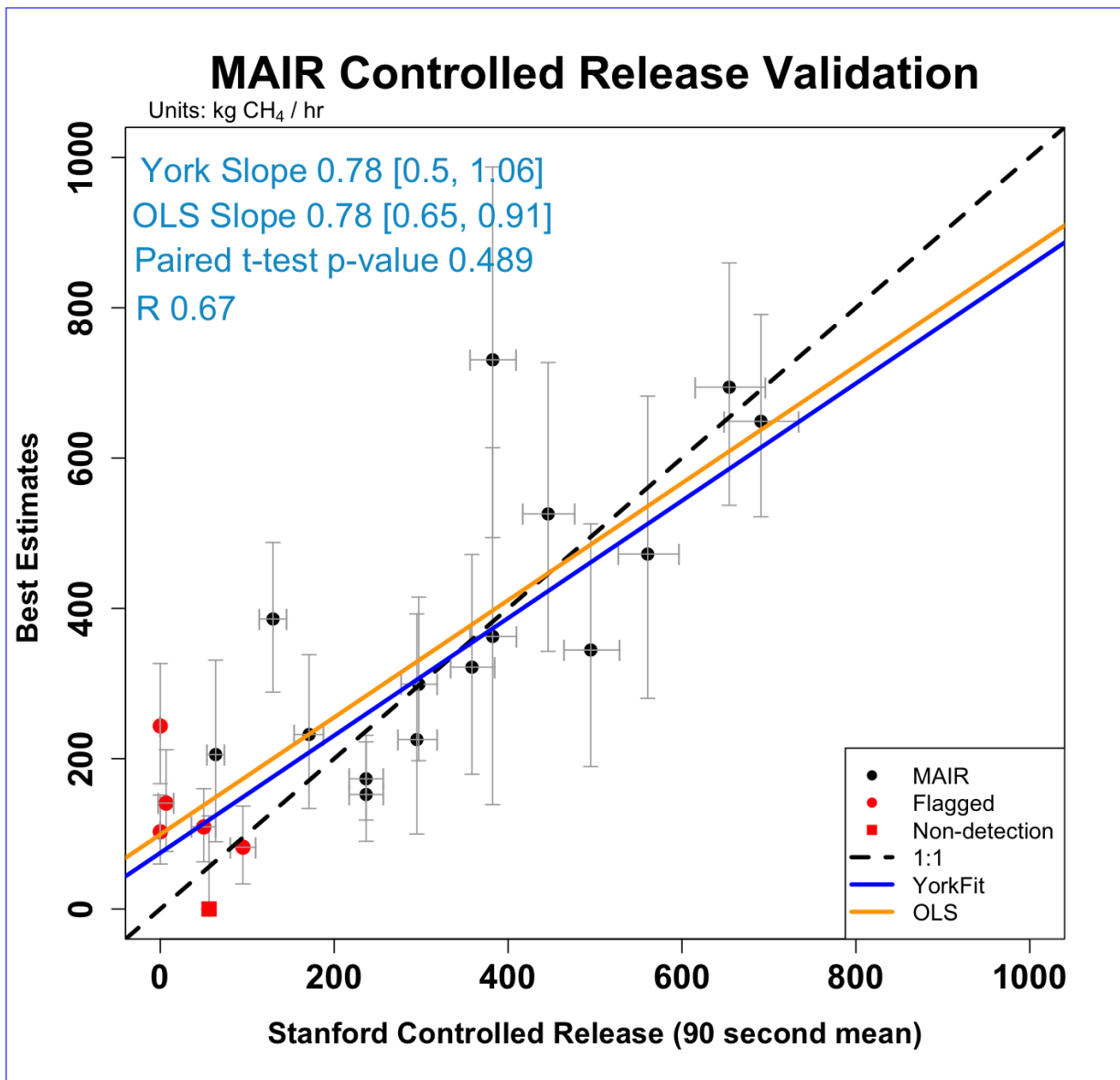



Figure S3: An analogous plot of Figure 3. The results presented in this figure are based on the "best estimates" reported on 1 February 2022. The black circles represent the post-unblinding comparison between the reported and estimated emissions. The red square represents the original comparison, which was later identified as no detection and removed. The red circles represent the flagged data points determined by the decision tree (S10). The blue solid line is the post-unblinding York fit. The orange is the Ordinary Least Squares (OLS) fit. ~~This figure was~~ The comparisons shown by the York and OLS fits are notional and do not ~~included~~ actually pertain to the information exchanged in the ~~original~~ unblinding ~~because~~. The errors in the MethaneAIR data had not yet been assessed in detail and, in fact, were overestimated. The MethaneAIR team did not ~~have~~ know the metered rates and the associated uncertainties were not documented.

S2.2   Unlit Flare

As shown in Figure S5, the flare emission rates range from 500 kg/hr to 2000 kg/hr.

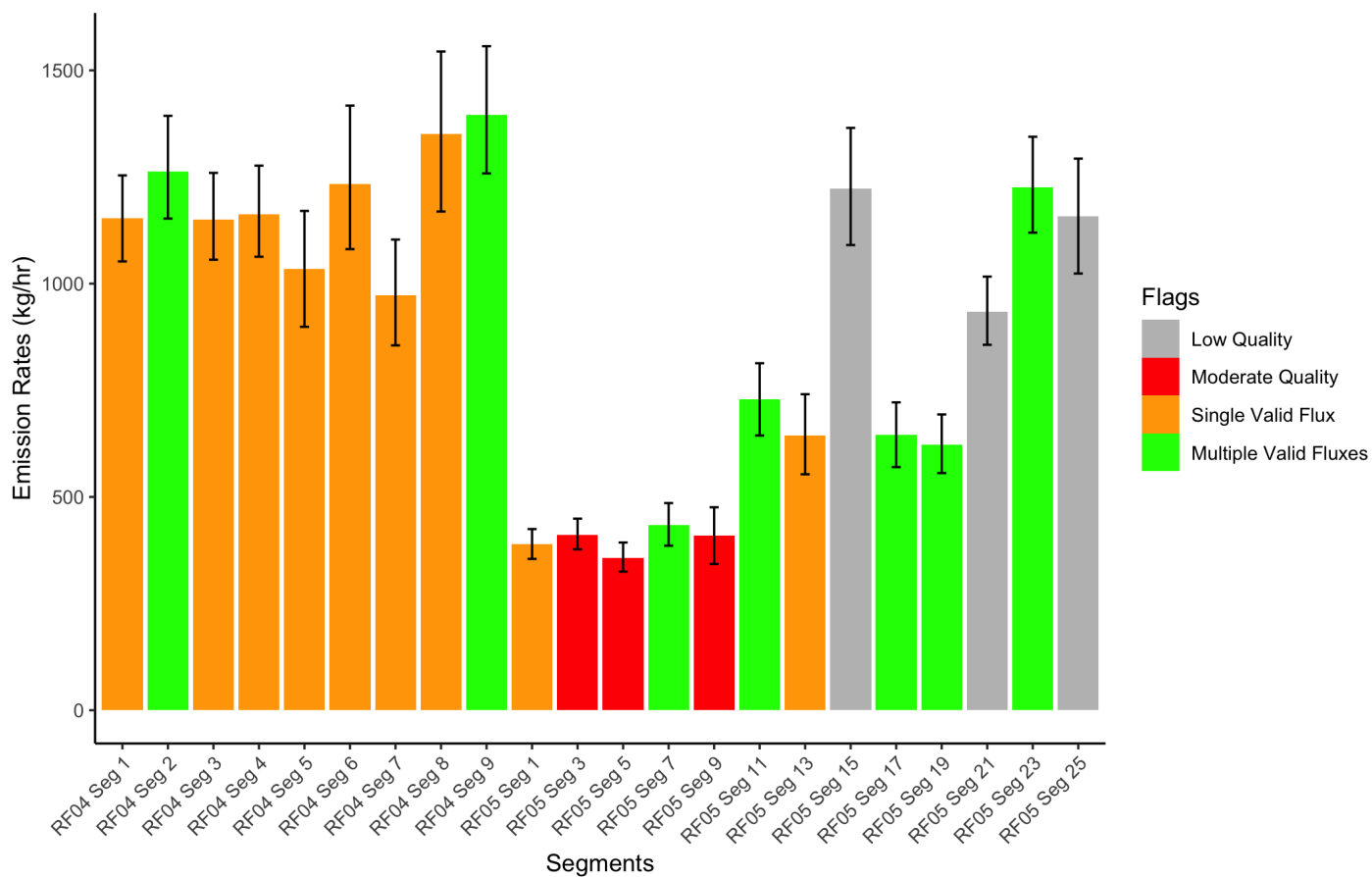



Figure S4: Methane emission estimates from an unlit flare during RF04 and RF05 using the mIME method. The endpoints are tagged and color-coded according to the Decision Tree in section S10.

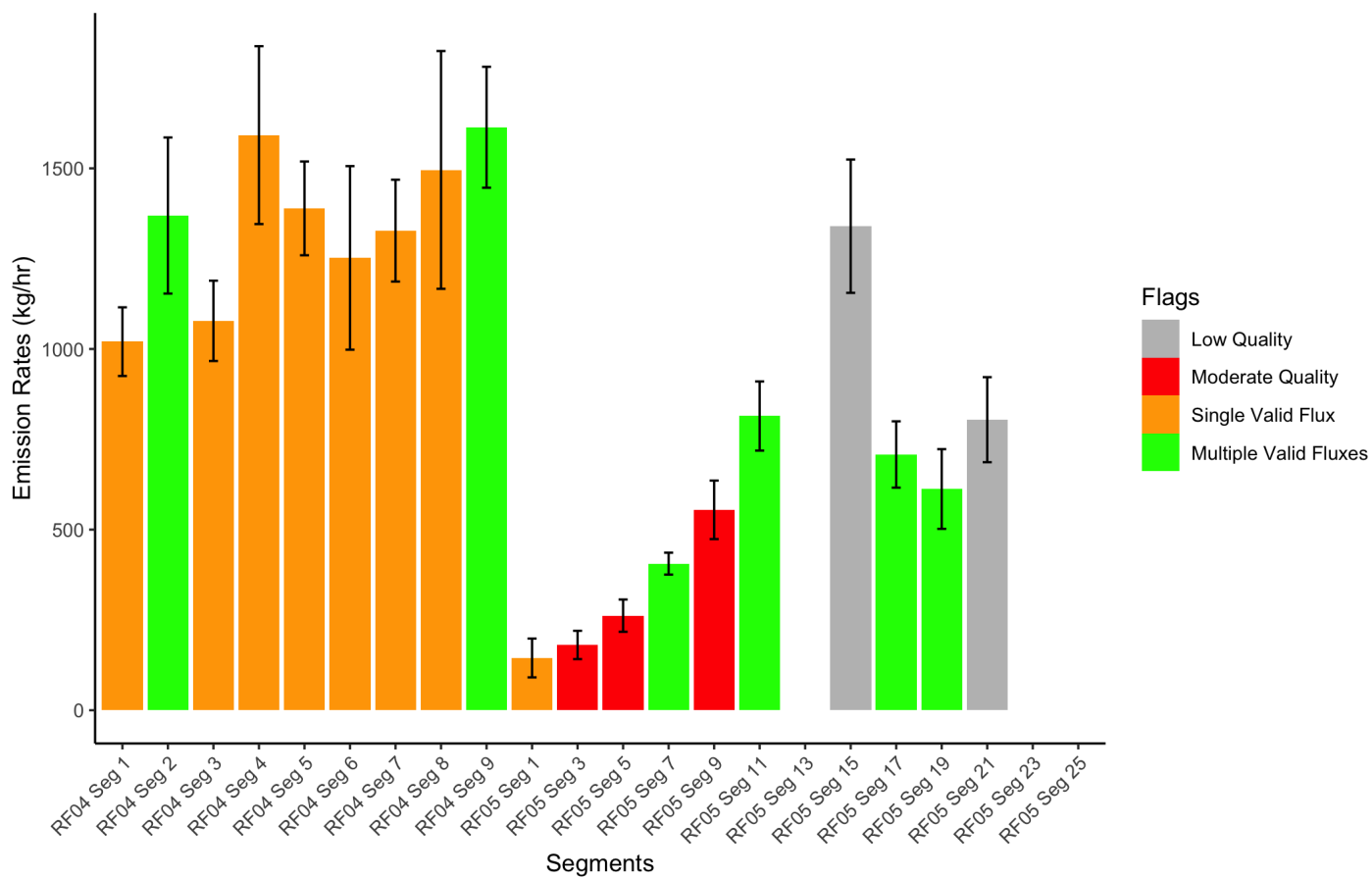



Figure S5: Methane emission estimates from an unlit flare during RF04 and RF05 using the DI method. The endpoints are tagged and color-coded according to the Decision Tree in section S10.

S2.3 MVD & Pipeline

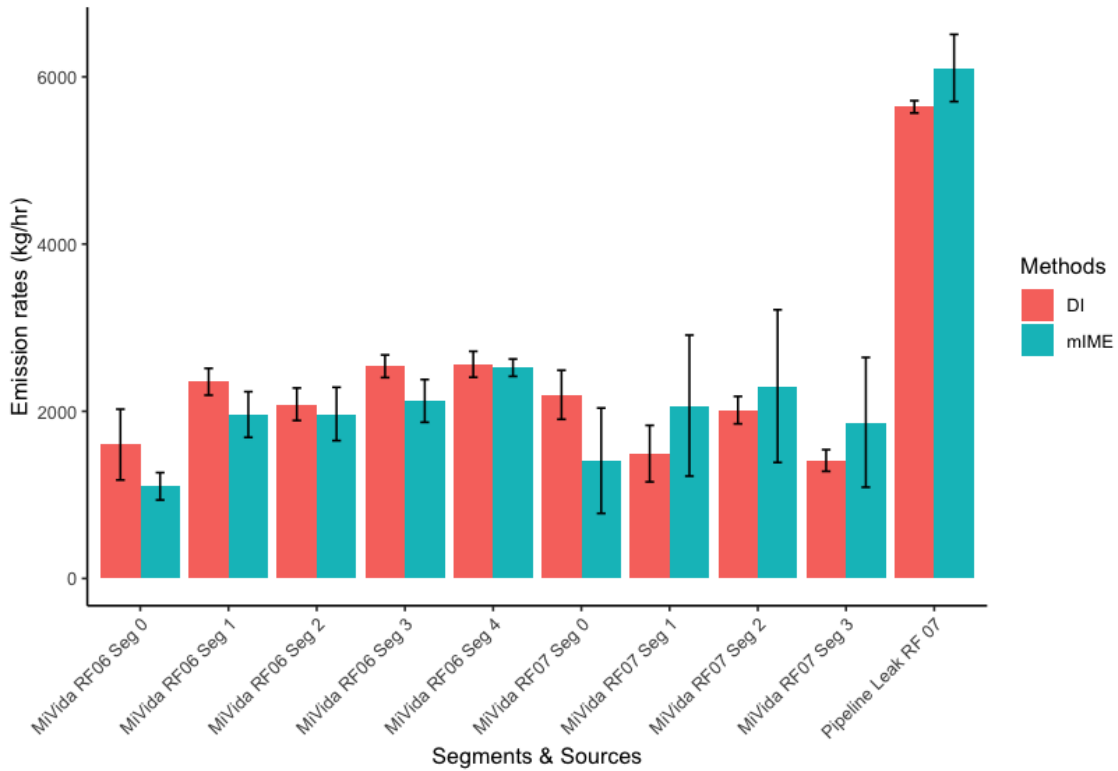



Figure S6: Barplot of the miscellaneous sources quantified in the Delaware Basin.

S3 Analysis Methods

S3.1 Wind Analysis

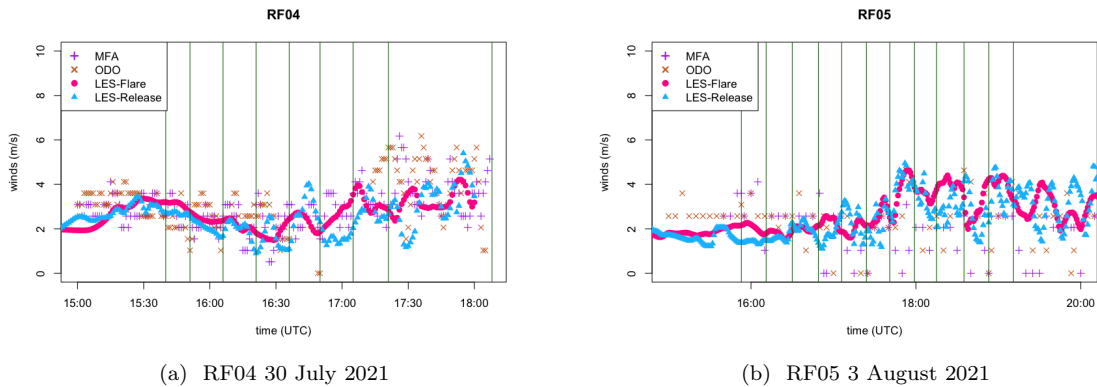
(a) RF04 30 July 2021

(b) RF05 3 August 2021

Figure S7: ASOS surface wind comparison from the blinded-volume controlled release experiments during (a) RF04 on 30 July 2021 and (b) RF05 on 3 August 2021. The purple crosses represent ASOS winds from the Midland International Air & Space Port (MAF), which is ~15 km from the blinded-volume controlled release site. The orange x's represent ASOS winds from the Odessa-schlemeyer Field Airport (ODO), which is ~17 km from the blinded-volume controlled release site. The magenta circles represent surface winds from the LES at the flare. The blue triangles represent surface winds from the LES at the blinded-volume controlled release site. The dark green vertical lines ~~represents~~ represent times when MAIR flew over the targets. All LES winds during the controlled release experiments and unlit flare observation agree with the ASOS winds (two-sided t-test p-values are greater than 0.05).

S3.2 Wind Product Analysis

To assess our LES winds compared to mesoscale HRRR winds and winds observed at the release site, we plot the estimated emission rates against the known emission rates from the Stanford team using different wind products. As shown in Figure S8, effective wind from the LES ~~provide~~ provides the highest R^2 value and slope closest to 1. Note that the choice of driven wind product is not limited to HRRR data; any wind products that can drive WRF can run WRF-LES. Although HRRR is the highest temporal and spatial resolution wind product for the Contiguous United States (CONUS), further study of other wind products outside CONUS is needed to determine the most appropriate wind products.

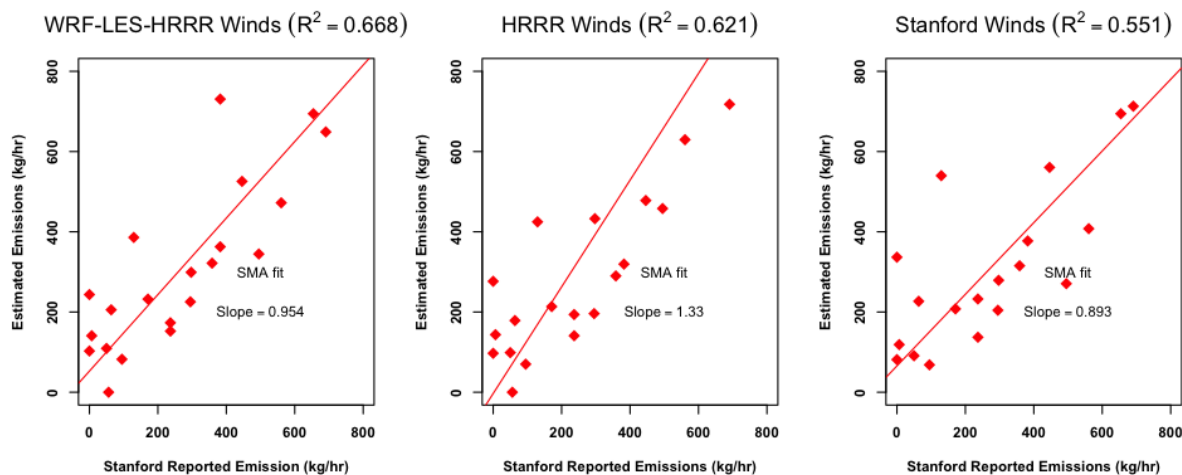



Figure S8: Estimated emission rates with the mIME using different wind products vs. the known emission rates from the Stanford team. From left to right: estimated emission rates based on effective winds from WRF-LES-HRRR, HRRR, and measured winds. The paired t-test between the estimated emissions and Stanford reported emissions suggests that we cannot reject the null hypothesis that a true difference in means is equal to 0 within 95% confidence interval for the estimated emissions based on WRF-LES-HRRR and Stanford observed winds with p-values of 0.166 and 0.309 respectively. However, the test suggests the alternative hypothesis of a true difference in means between the emissions derived from HRRR winds and the reported emissions with a p-value of 0.0408.

S3.3 Threshold Analysis

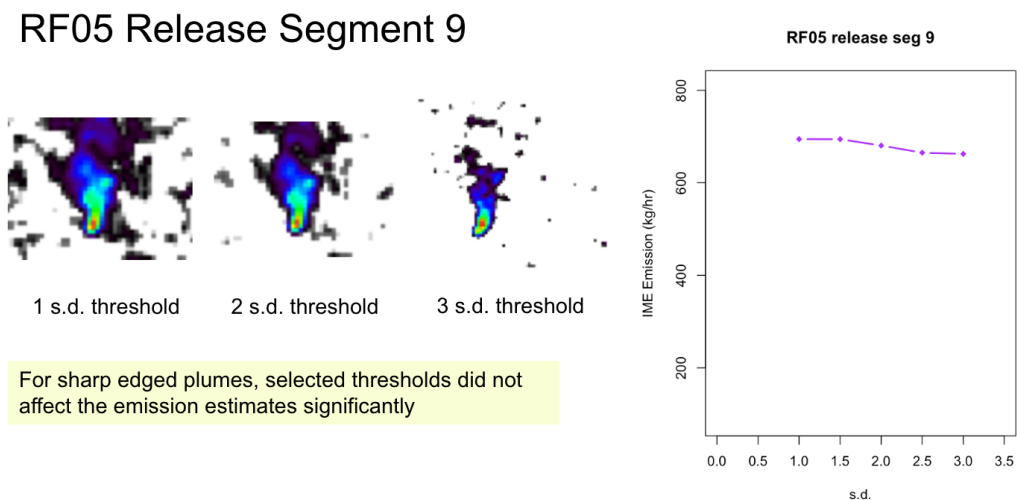



Figure S9: The thresholds do not affect the IME values of the plume of interest.

S3.4 Decision Tree

To understand the relationship between the actual emission rates and the emission estimates, we attempted to use several regression models for our data. York regression was the most appropriate one because York includes correlated errors from both axes. OLS was included in this study since OLS is often used in this type of experiment. See Sherwin et al. (2022), SI Section S4, for further discussion of tradeoffs associated with York and OLS fits in single-blind tests of methane remote sensing.

Before the actual blinded-volume controlled release values were revealed, we developed a decision tree workflow that assigns a confidence flag to each case (Figure S10). The original version of our decision tree was developed purely based on our analysis of our measurement system and algorithms. It was then adjusted slightly after unblinding the controlled release metered emissions. When the signal-to-noise ratio (SNR) for the plume, relative to the background, is below 1 or significant interference is observed, we classify the case as either no detection (Endpoint 0) or poor quality (Endpoint 1). When the SNR is between 1 and 2.5, the estimates are near the detection limit, providing a moderate-quality estimate (Endpoint 2). Only the mIME method can be used, and the confidence will likely be low. When the SNR is above 2.5, we checked if the plume size is greater than one eddy. If the plumes are smaller than one eddy, only the mIME method can be used (Endpoint 3). If the plumes are larger than one eddy, all three methods developed were applied (Endpoint 4). For the last step, the WRF-HRRR-LES were compared to the nearby ASOS winds and later assigned quality flags based on agreement to the observed winds. If the winds disagree (two-sided t-test p-value is less than 0.05), the wind quality is low. When winds agree (two-sided t-test p-value is greater than 0.05), the wind quality is high. All LES winds during the controlled release experiments and unlit flare observation agree within tolerance with the ASOS winds.

As expected, the DI and the Ratio approaches work well with larger sources such as the unlit flare, MiVida Gas Processing Plant, or the pipeline leak plume. However, the ratio approach is limited by the resolution mismatch with the sensor and the LES domain size, making this approach less favorable. So, we decided to proceed with the mIME and the DI methods.

After the metered blinded-volume controlled release values were revealed, the decision tree was revised to include additional steps that help remove false positive cases and help us determine the detection limits of MethaneAIR.

By filtering out cases that (i) have the absolute differences between mIME emission estimates using a threshold of 1 and 2 standard deviations divided by emission estimate using a threshold of 1 standard deviation greater than 0.5 (i.e., 2 s.d. case has a much different plume shape) or (ii) no emission estimates at the threshold of 2 standard deviations, we designate as "below detection limit" 6 cases from the blinded-volume controlled release experiment. One was a false negative case, where methane enhancements were not detected. The other four were emission estimates around the detection limit of 200 kg/hr. The part of the decision tree added after unblinding, intended to reduce false positives in the contaminated environment of this test, is boxed in red in Figure S10.

MethaneAIR/SAT Plume Quantification Decision Tree

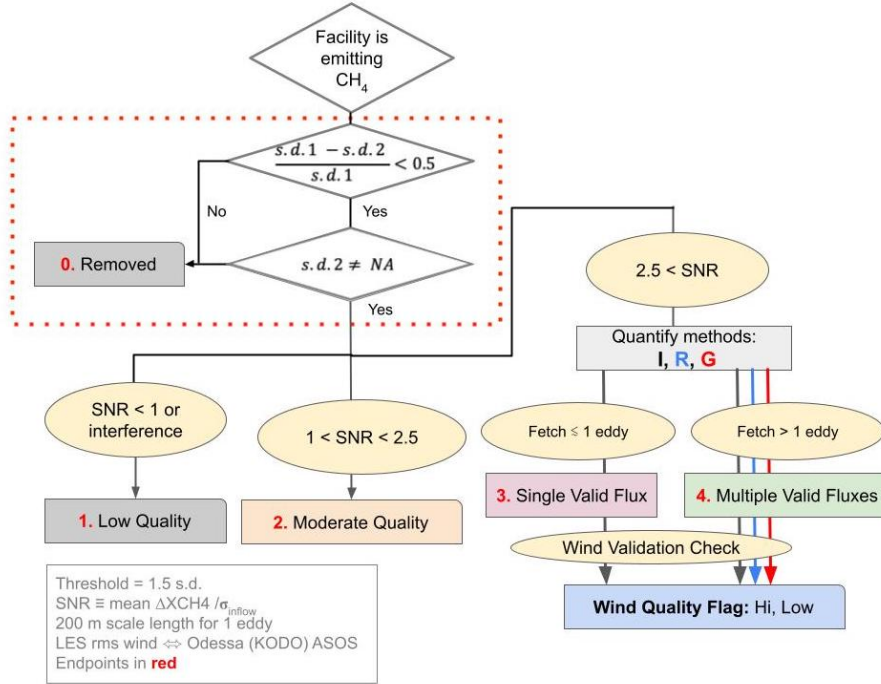



Figure S10: Updated decision tree classifying quality levels of emission estimates. Most of the steps were developed purely based on our analysis of the data and signal processing. The steps in the red dashed box were added after the unblinding process. First, we filter out all the estimates that (i) have the absolute differences between mIME emission estimates using ~~threshold~~ thresholds of 1 and 2 standard deviations divided by emission estimate using a threshold of 1 standard deviation to be greater than 0.5 or (ii) no emission estimates at the threshold of 2 standard deviations. After that, for each estimate with SNR below 1 or significant interference was observed, we classify the case as either no detection (~~End point~~ Endpoint 0) or poor quality (~~End point~~ Endpoint 1). ~~When~~ The estimates are moderate quality when the SNR is between 1 and 2.5 ~~, the estimates are moderate quality~~ (~~End point~~ Endpoint 2). Only the mIME method can be used. When the SNR is above 2.5, we checked if the ~~size of the~~ plume size is greater than one eddy. If the plumes are smaller than one eddy, only the mIME method can be used (~~End point~~ Endpoint 3). ~~If~~ All three different methods developed were applied if the plumes are larger than one eddy ~~, all three different methods developed were applied~~ (~~End point~~ Endpoint 4). For the last step, the WRF-HRRR-LES were compared to the nearby ASOS winds and later assigned quality flags based on agreement to the observed winds. If the winds disagree, the wind quality is low. When winds agree, the wind quality is high.

S3.5   Additional Attempts to Determine Contamination Near the Methane Plumes

Although our decision tree (S3.4) removed ~~6~~ six plumes (below detection) from the total 21 plumes based on the ~~signal to noise~~ signal-to-noise ratio and the size of the plumes, we wanted to investigate the limitations of our sensors and methods further. We decided to perform two more analyses with the hypothesis that these two will help us better understand the limitations.

The first analysis was to perform a k-s test between XCH_4 in the inflow region of each scene and a clear scene defined as latitude 32.10 - 32.11 and longitude -102.32 to -102.31. When the p-values of the k-s test ~~is~~ are less than 0.05, we can reject the null hypothesis that the XCH_4 observed in the inflow and clear scene came from the same distribution. A positive outcome from the K-S test suggests that there may be extraneous methane contamination in the inflow.

The second analysis was to filter the clumps with 60 pixels and 100 pixels. The new choices of thresholds are arbitrarily rounded to the multiples of ~~original 20 pixel~~ the original 20-pixel threshold and fractions of a large

eddy (~~approximate~~ approximately 100 m). The original ~~20 pixel~~ 20-pixel threshold or 10 m × 10 m × 20 pixels is equivalent to a 45 m × 45 m plume, which is approximately half of the size of an eddy. The ~~60 pixel~~ 60-pixel threshold is equivalent to a 77 m × 77 m plume, which is approximately ~~two third~~ two-thirds of the size of an eddy. The ~~100 pixel~~ 100-pixel threshold is equivalent to a 100 m × 100 m plume, which is approximately the size of an eddy.

Table S6: Summary table of removed (not detected) RF04 plumes based on three methods: decision tree (S3.4), 60 pixel threshold, and K-S test.

Segment	Decision Tree	60 pixels	K-S test
1			x
2			x
3	x	x	
4	x	x	
5		x	x
6			
7	x	x	
8			
9	x	x	

Table S7: Summary table of removed RF05 plumes based on three methods: decision tree (S3.4), 60 pixel threshold, and K-S test.

Segment	Decision Tree	60 pixels	K-S test
1	x	x	x
3			x
5	x	x	x
7	x	x	
9		x	x
11			
13	x	x	x
15			
17	x	x	
21			
23			

S4   Cropped Blinded-Volume controlled release Plumes

The following figures are cropped blinded-volume controlled release plumes from ~~both~~ RF04 and RF05. The plumes were identified based on the known location of the blinded-volume controlled release site (grey triangles). To reduce the noise of the raw 1×1 Level 3 data, the data were later filtered by a Gaussian filter (~~7 pixel x 7 pixel~~ 7-pixel x 7-pixel kernel with the full width at ~~hald~~ a maximum of 2 pixels). We then identified the inflow region and removed the pixels below the 1.5 standard deviation of the inflow threshold. After that, we removed the small clumps with fewer than 20 pixels from the scene of interest.

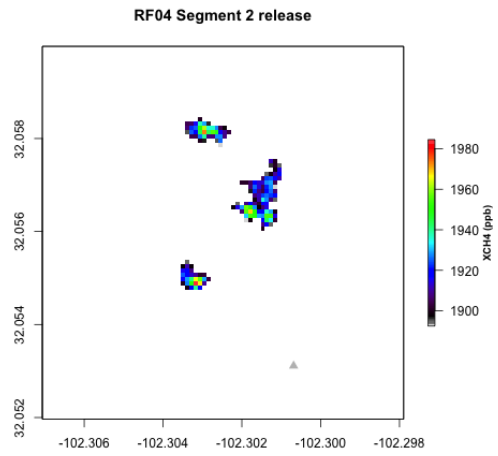
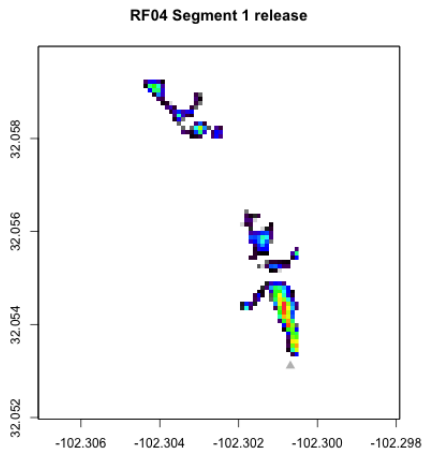



Figure S11: Filtered blinded-volume controlled release RF04 Segment 1.

Figure S12: Filtered blinded-volume controlled release RF04 Segment 2.

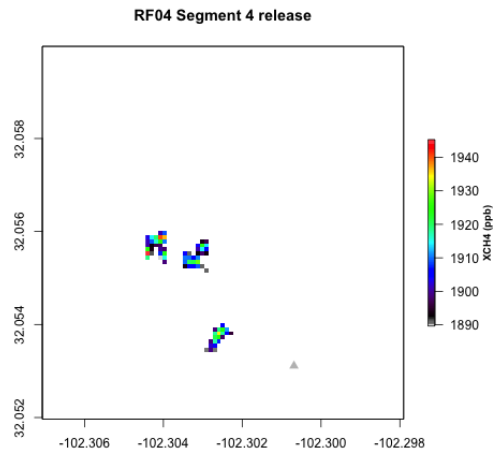
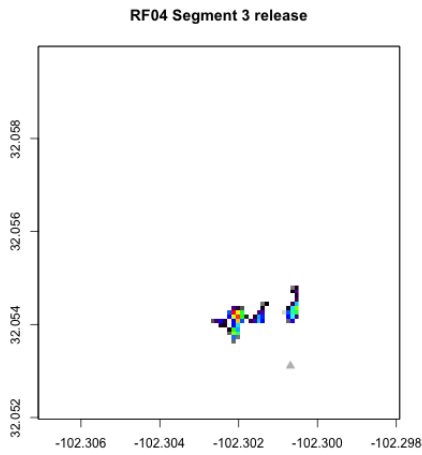



Figure S13: Filtered blinded-volume controlled release RF04 Segment 3.

Figure S14: Filtered blinded-volume controlled release RF04 Segment 4.

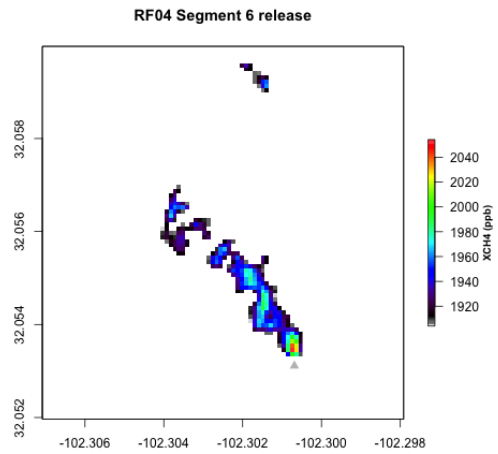
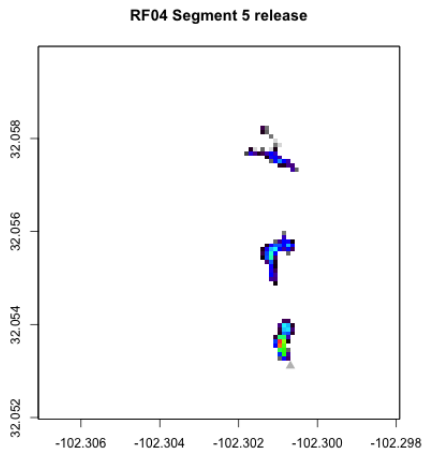



Figure S15: Filtered blinded-volume controlled release RF04 Segment 5.

Figure S16: Filtered blinded-volume controlled release RF04 Segment 6.

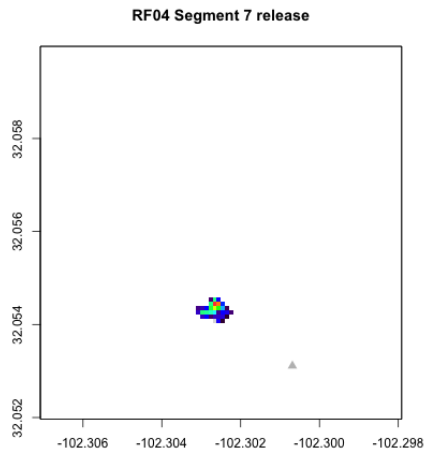



Figure S17: Filtered blinded-volume controlled release RF04 Segment 7.

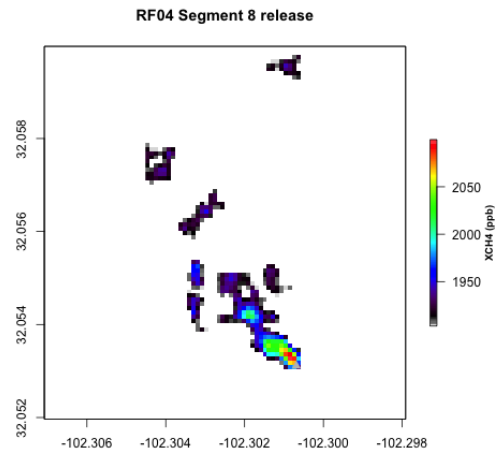



Figure S18: Filtered blinded-volume controlled release RF04 Segment 8.

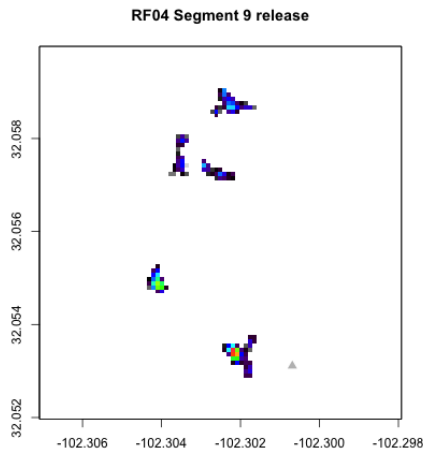



Figure S19: Filtered blinded-volume controlled release RF04 Segment 9.

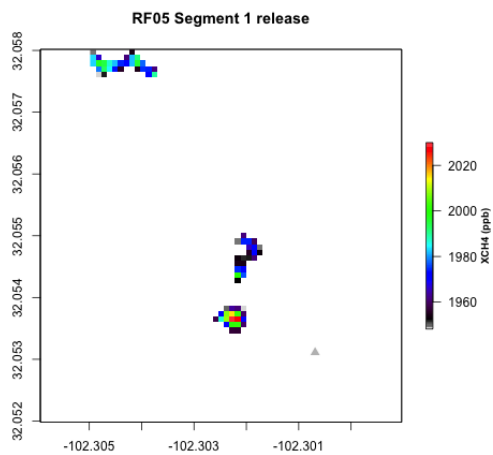



Figure S20: Filtered blinded-volume controlled release RF05 Segment 1.

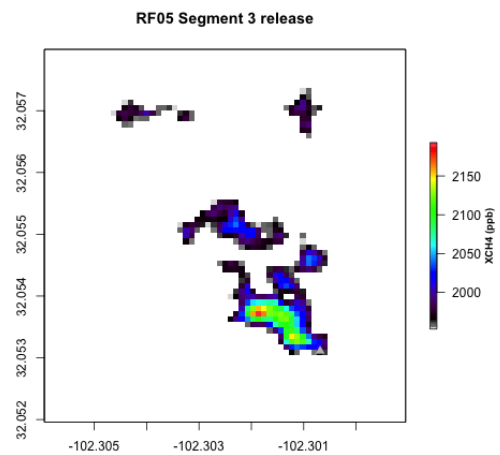



Figure S21: Filtered blinded-volume controlled release RF05 Segment 3.

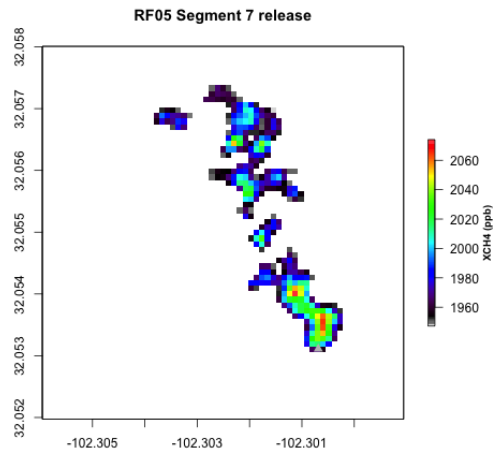
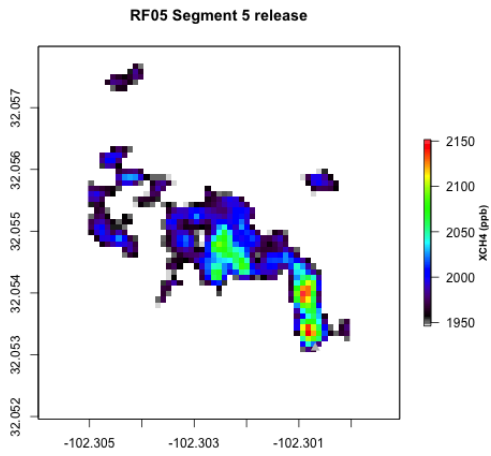



Figure S22: Filtered blinded-volume controlled release RF05 Segment 5.   Figure S23: Filtered blinded-volume controlled release RF05 Segment 7.

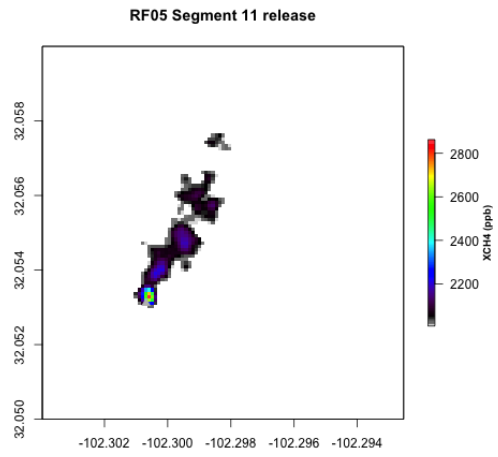
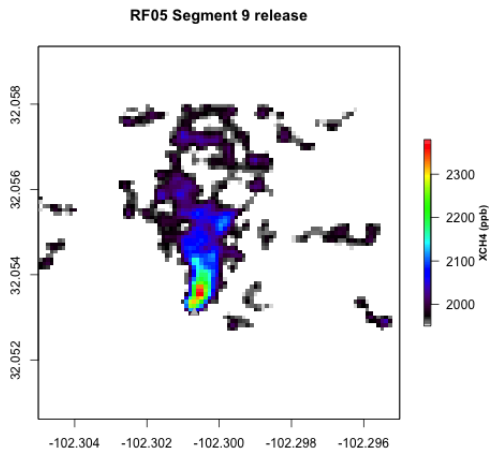



Figure S24: Filtered blinded-volume controlled release RF05 Segment 9.   Figure S25: Filtered blinded-volume controlled release RF05 Segment 11.

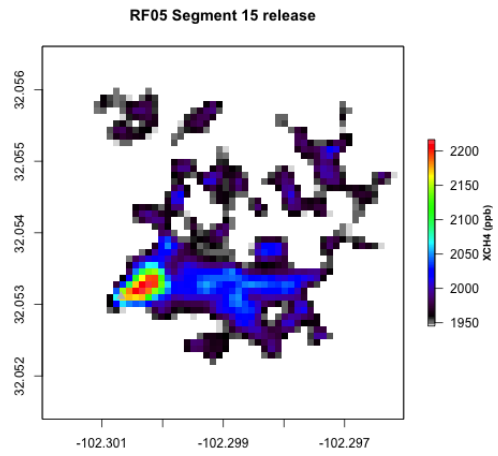
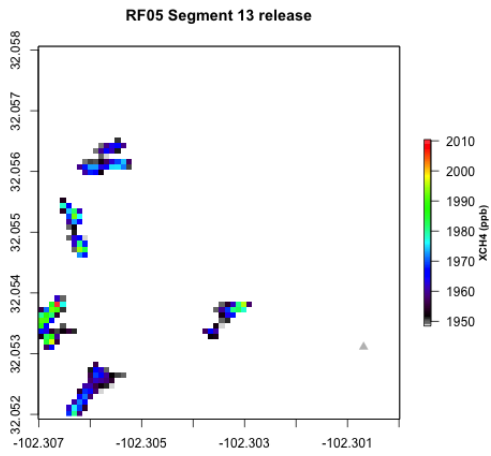



Figure S26: Filtered blinded-volume controlled release RF05 Segment 13.   Figure S27: Filtered blinded-volume controlled release RF05 Segment 15.

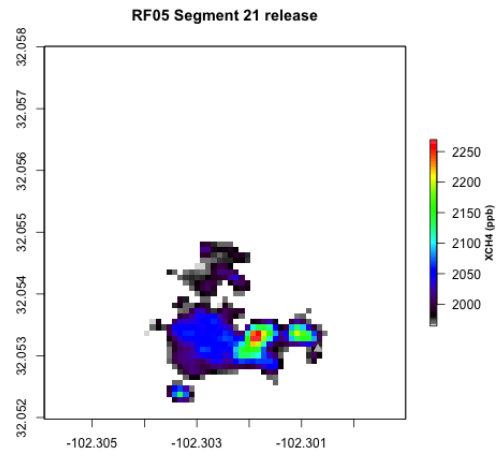
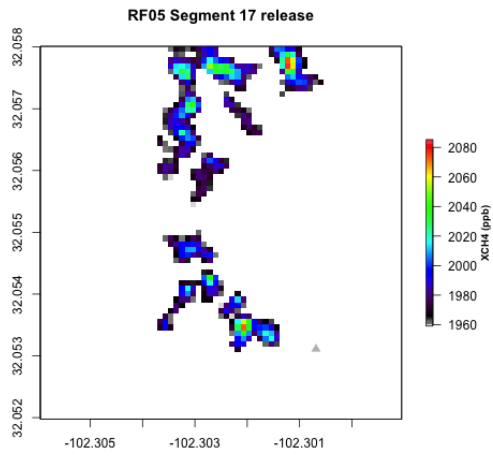



Figure S28: Filtered blinded-volume controlled release Figure S29: Filtered blinded-volume controlled release
RF05 Segment 17.   RF05 Segment 21.

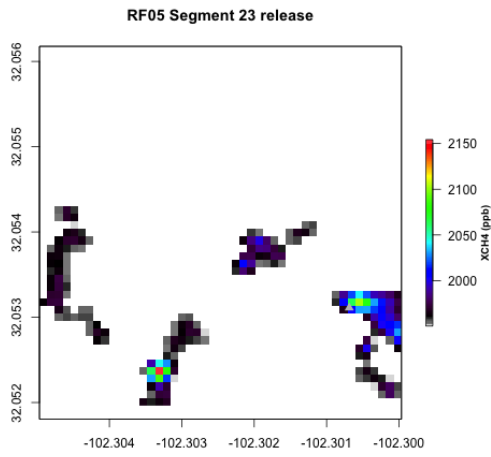



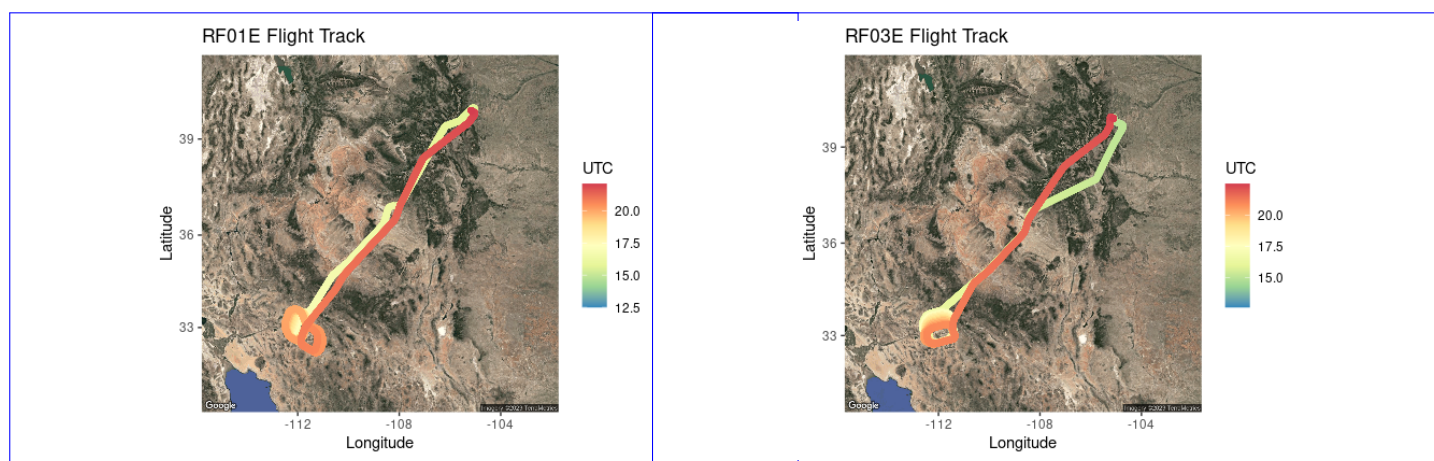
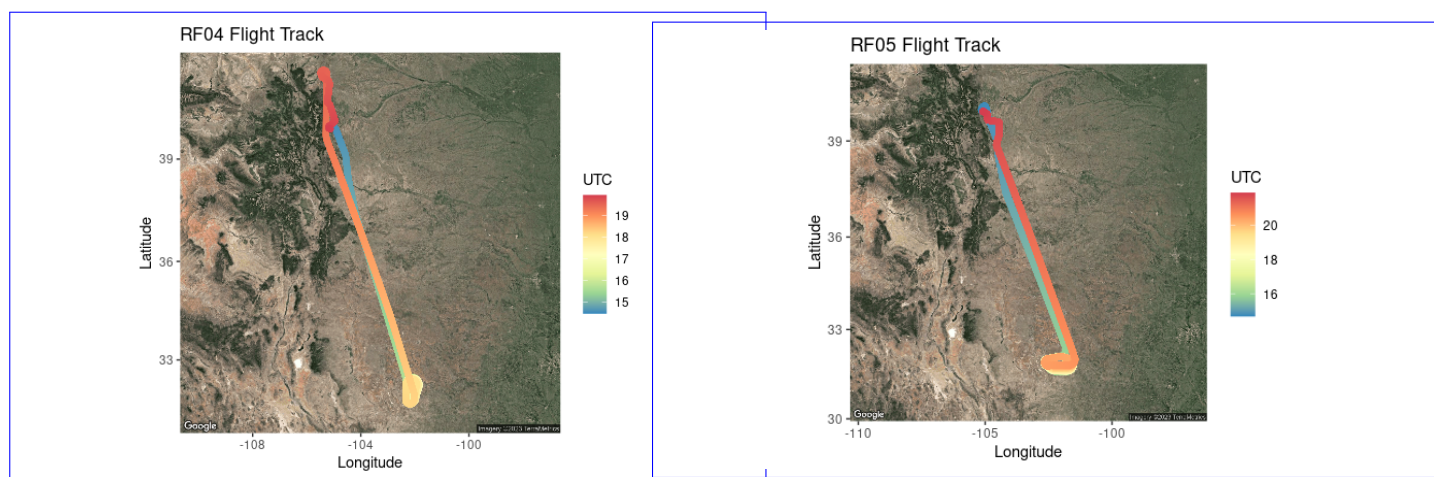
Figure S30: Filtered blinded-volume controlled release
RF05 Segment 23.

S5   MethaneAIR flight tracks

Figure S31: RF01E

Figure S32: RF03E

Figure S33: RF04

Figure S34: RF05

References

Conley, S., Faloona, I., Mehrotra, S., Suard, M., Lenschow, D. H., Sweeney, C., Herndon, S., Schwietzke, S., Pétron, G., Pifer, J., Kort, E. A., and Schnell, R.: Application of Gauss's theorem to quantify localized surface emissions from airborne measurements of wind and trace gases, Atmospheric Measurement Techniques, 10, 3345–3358, https://doi.org/10.5194/amt-10-3345-2017, 2017.

Irakulis-Loitxate, I., Guanter, L., Liu, Y.-N., Varon, D. J., Maasakkers, J. D., Zhang, Y., Chulakadabba, A., Wofsy, S. C., Thorpe, A. K., Duren, R. M., Frankenberg, C., Lyon, D. R., Hmiel, B., Cusworth, D. H., Zhang, Y., Segl, K., Gorroño, J., Sánchez-García, E., Sulprizio, M. P., Cao, K., Zhu, H., Liang, J., Li, X., Aben, I., and Jacob, D. J.: Satellite-based survey of extreme methane emissions in the Permian basin, Science Advances, 7, eabf4507, https://doi.org/10.1126/sciadv.abf4507, 2021.

Rutherford, J. S., Sherwin, E. D., Chen, Y., Aminfard, S., and Brandt, A. R.: Evaluating methane emission quantification performance and uncertainty of aerial technologies via high-volume single-blind controlled releases, URL https://eartharxiv.org/repository/view/5113/, publisher: EarthArXiv, 2023.

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 methane emissions detection and quantification, URL https://eartharxiv.org/repository/view/3465/, publisher: EarthArXiv, 2022.

Sun, K., Zhu, L., Cady-Pereira, K., Chan Miller, C., Chance, K., Clarisse, L., Coheur, P.-F., González Abad, G., Huang, G., Liu, X., Van Damme, M., Yang, K., and Zondlo, M.: A physics-based approach to oversample

multi-satellite, multispecies observations to a common grid, Atmospheric Measurement Techniques, 11, 6679–6701, https://doi.org/10.5194/amt-11-6679-2018, publisher: Copernicus GmbH, 2018.

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, Atmospheric Measurement Techniques, 11, 5673–5686, https://doi.org/10.5194/amt-11-5673-2018, 2018.

York, D.: Least squares fitting of a straight line with correlated errors, Earth and Planetary Science Letters, 5, 320–324, https://doi.org/10.1016/S0012-821X(68)80059-7, 1968.

York, D., Evensen, N. M., Martınez, M. L., and De Basabe Delgado, J.: Unified equations for the slope, intercept, and standard errors of the best straight line, American Journal of Physics, 72, 367–375, https://doi.org/10.1119/1.1632486, publisher: American Association of Physics Teachers, 2004.