Sectoral contributions of high-emitting methane point sources from major U.S. onshore oil and gas producing basins using airborne measurements from MethaneAIR

Jack D. Warren[*1], Maryann Sargent[2], James P. Williams[1], Mark Omara[1,3], Christopher C. Miller[1,2,3], Sebastien Roche[1,2,3], Katlyn Mackay[1], Ethan Manninen[2], Apisada Chulakadabba[2], Anthony Himmelberger[3], Joshua Benmergui[1,2,3], Zhan Zhang[2], Luis Guanter[1], Steven Wofsy[2], Ritesh Gautam*[1,3]

[1]Environmental Defense Fund, New York, NY, USA 10010

[2]Harvard University, Cambridge, MA, USA 02138

[3]MethaneSAT, LLC, Austin, TX, USA 78701

Correspondence to: Jack D. Warren (jwarren@edf.org), Ritesh Gautam (rgautam@edf.org)

Supplemental Information

S1. Thresholding method used to identify methane plumes in MethaneAIR scenes.

The gridded flux map used to identify methane plumes is shown in Figure S1.1. The left panel shows the 600m x 600m boxes being tiled across the XCH4 image. This size box was chosen as it is a typical plume width that we observed, so the boxes are on the scale of a plume, and can fully enclose a plume's origin. The DI is calculated around each box using HRRR winds to estimate the flux from within the box. The boxes are oversampled (not shown in the figure), moving the box over by 200m in each step, to produce a gridded flux product on a 200m x 200m grid (right panel). Note that while the methane plume in the left panel extends across the length of the figure, in the right panel the upwind end near the source produces much higher (and lower) flux values than the downwind end. While the gridded flux product was quite useful on the plume scale, we found that summing totals across the scene was not accurate at estimating regional emissions, as some assumptions in the method, such as no significant mass loss through the top and bottom of the box, were not valid in the large scale, but were valid on the scale of individual plumes. We use HRRR winds at level 3, ~80m, which is the typical mixing height of a plume on length scales of 100-2000 m from the source, which is the scale at which we typically observe plumes with MAIR.

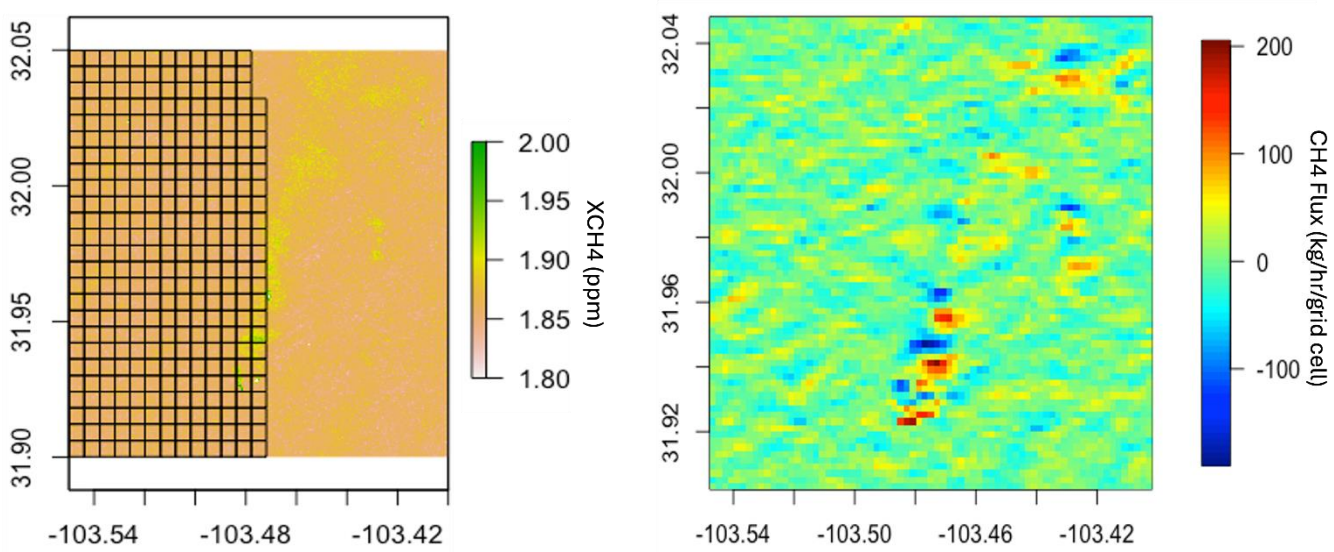


Figure S1.1. Left: 600 m grid used to calculate divergence integral (DI) for a small area of flight RF06. DI is calculated to find the flux through each box. Oversampling is used by moving sequentially moving the 600m box over by 200m across the scene, producing the 200m gridded image to the right. Right: Gridded DI flux product corresponding to MethaneAIR scene on the right.

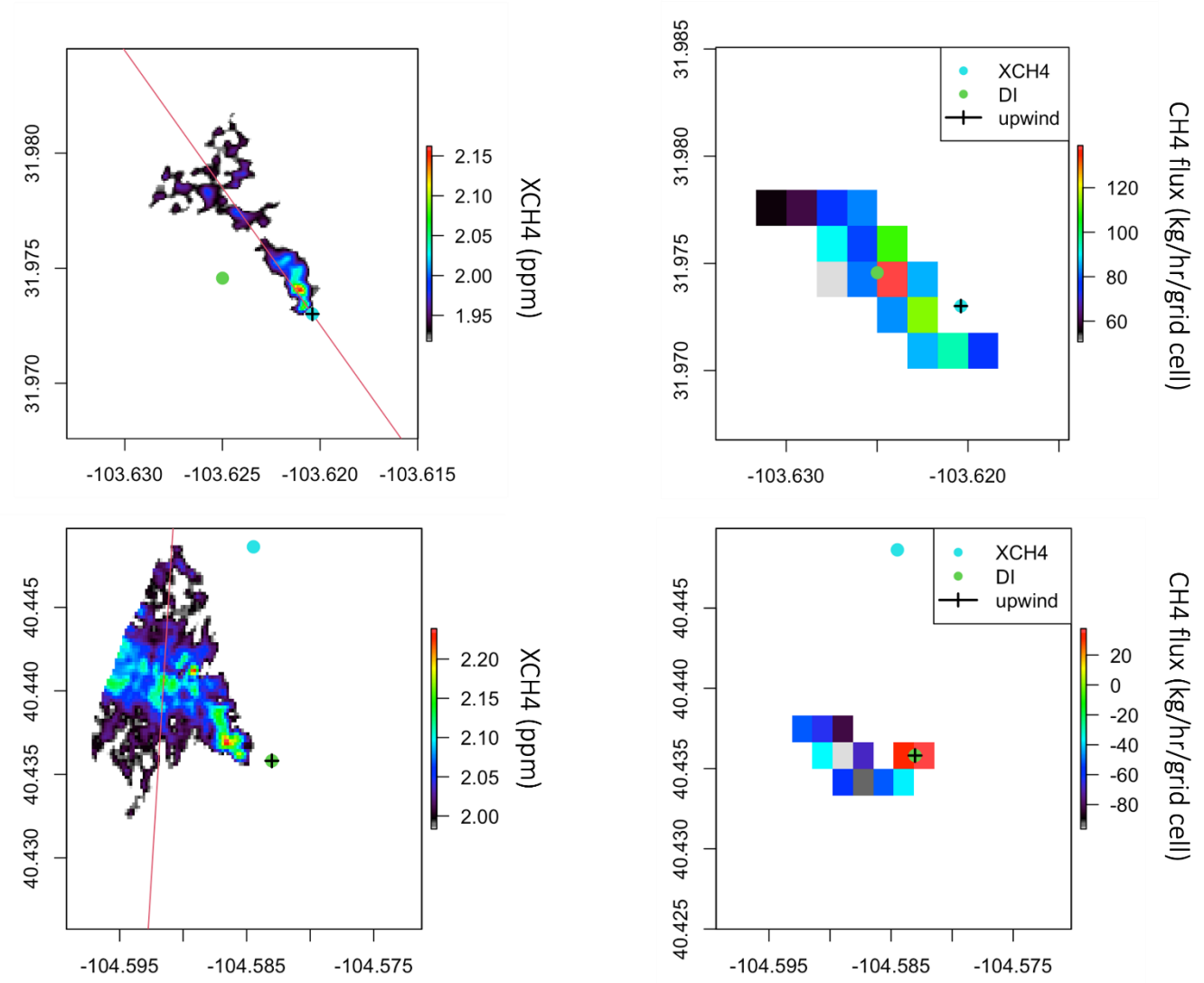


Figure S1.2: Left column: Masked XCH4 plumes. Right column: Masked DI flux "clumps". Red line shows the major axis of the XCH4 plume. The top and bottom rows show 2 different plumes. The filled cyan circle is the upwind end of the plume; the open cyan circle is the downwind end of the plume. The filled blue circle is the center of the positive portion of the DI flux "clump". The plus is the point chosen as the origin of the plume, whichever is upwind between the upwind end of the XCH4 plume and the center of the DI flux clump.

The filled blue circle is the center of the positive portion of the DI flux "clump". The plus is the point chosen as the origin of the plume, whichever is upwind between the upwind end of the XCH4 plume and the center of the DI flux clump. The top row is a typical plume form, where the upwind end of the plume is correctly found to be the plume origin. The bottom row

shows a ”pancake” shaped plume, which can happen at low wind speeds, where the long axis of the plume is actually perpendicular to the wind direction. In this case, the center of the DI flux ”clump” is chosen as the upwind end.

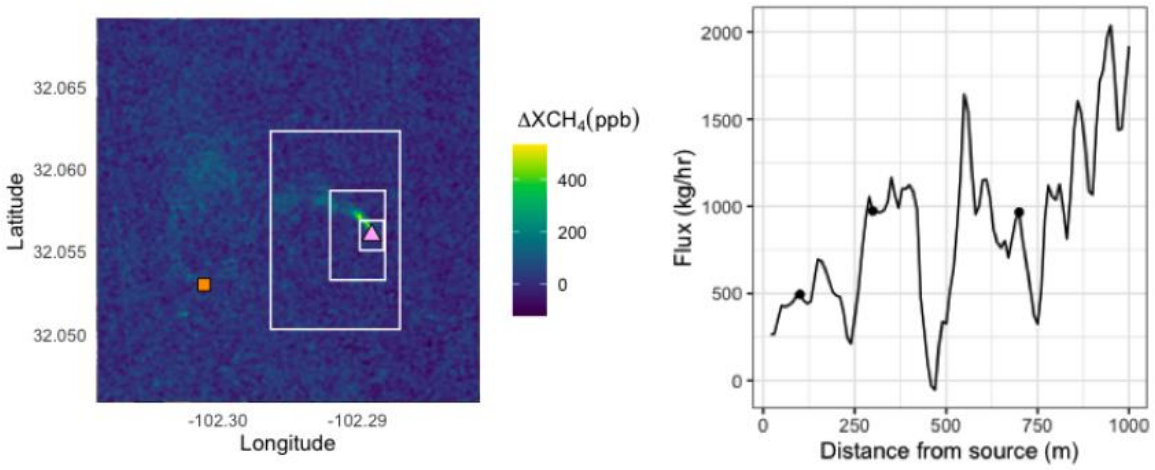


Figure S1.3: From Chulakadabba et al., 2023, Figure 2. MethaneAIR image of a single-blind volume-controlled release (orange square) and an unlit flare (pink triangle) observed on 3 August 2021. Left: Three example rectangles are used to calculate the flux divergence. Right: Calculated flux divergence 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 left panel. 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 250m 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 dm/dt. In this example, we averaged the DI from 80–700m to avoid influence from other sources nearby that increase the DI beyond 700m from the source.

S2. Facility-level attribution and repeat detection identification methods

Attribution of point sources to facility types has been achieved through spatially querying known infrastructure locations (Hmiel et al. 2022), or by manually review of available imagery (Cusworth et al 2022). Here we apply both methods. First, we assembled a collection of public infrastructure databases, including state and federal level inventories of air emission sources(Active Air Pollution Emitting Facilities; New Mexico Environment Department Data; Ohi Emissions Inventory point source data; Coal Mining Operation - Underground Mines; West Virginia Department of Environmental Protection GIS Data; Underground and Surface Coal Mines), and O&G infrastructure collections (Omara et al. 2023) and several sources dedicated

to other methane emitting sectors (LMOP Landfill and Project Database; Coal Mining Operation - Underground Mines; CAFOs in the US; Mineral Resources of Wyoming).

Infrastructure databases typically represent facilities as point locations, so joining the origin of a plume must account for variability in where the single point representing the facility is located within the facilities footprint. For facilities with a small footprint, like O&G well sites, spatially joining the plume origin to the single known-infrastructure point is more straightforward compared to a large footprint facility, such as a processing plant. Conversely, the density of different facilities, such as a wellsite in close proximity to a compressor station, can pose a problem when using excessively large spatial buffers to query known locations and make a singular determination. To address both issues, we used a series of consecutive spatial buffers in combination with a prioritization hierarchy based on typical footprint sizes of the facilities to make a single attribution of facility type.

Features within this assembled infrastructure database were searched within successive distances of 150m, 300m, and 500m radii around all point source identifications. If no facility was found within a 150m radius, then the database was queried again at 300m and so on. If two or more features were identified within a given radius, a priority was given to the feature according to this facility type hierarchy corresponding to facility footprint size:

processing, CAFO > compressor station, waste, coal > tank battery, other > O&G wells

If multiple features were identified within the same level of the hierarchy, then the closest of these infrastructure points was used for the attribution. These attributions were then manually reviewed alongside available satellite or streetview imagery. For cases where a facility type could not be determined, such as where the plume was ambiguously between two facilities of different types, or no information nor recent imagery was available on the underlying location, the attribution was left as unknown. If the plume was between two facilities of the same sector during manual review, such as between an O&G well and a compressor station, then the sector was noted but specific attribution left as unknown. All point source detections were then spatially aggregated within a 300 m radius to identify both the number of unique locations where emissions were detected, and locations where emissions were detected multiple times across individual flights. Attributions across the aggregated detections were then manually reviewed for consistency to one another as an additional QAQC step for attribution and aggregation to a single facility.

Definitions for attributed facility types are as follows:

Compressor stations: facilities that use compressors to send produced gas along a pipeline network. While some production sites do have compressors on-site, these facilities do not have wellheads. A distinction between gathering and boosting segment stations, and transmission and storage segment stations is not made.

O&G wellsites: includes pumpjack-only sites and wellpads with other complex equipment present. While assembling infrastructure databases, it was noted that the delineation of a pad with a centralized tank battery, used for gathering production from multiple wellsites, was not consistent. In some regions or sources, centralized tank batteries were defined specifically while elsewhere they were referred to as just wellsites. Due to this inconsistency, we include centralized tank batteries in this category, despite the possibility that emissions characteristics may differ.

Pipelines: natural gas pipelines at least 150m from any pad. A distinction is not made between gathering and transmission pipelines.

Processing: facilities that ready natural gas for distribution by further separating liquids, heavier hydrocarbons, and other contaminants.

Waste: municipal and industrial waste landfills and their associated treatment facilities

Coal: Any facility related to coal mining, such as a mine entryway or vent.

CAFO: Concentrated animal feeding operations, such as for dairy or beef production.

Other: Any other non-O&G sector facilities, such as power plants.

S3. Sub-basin regions of interest

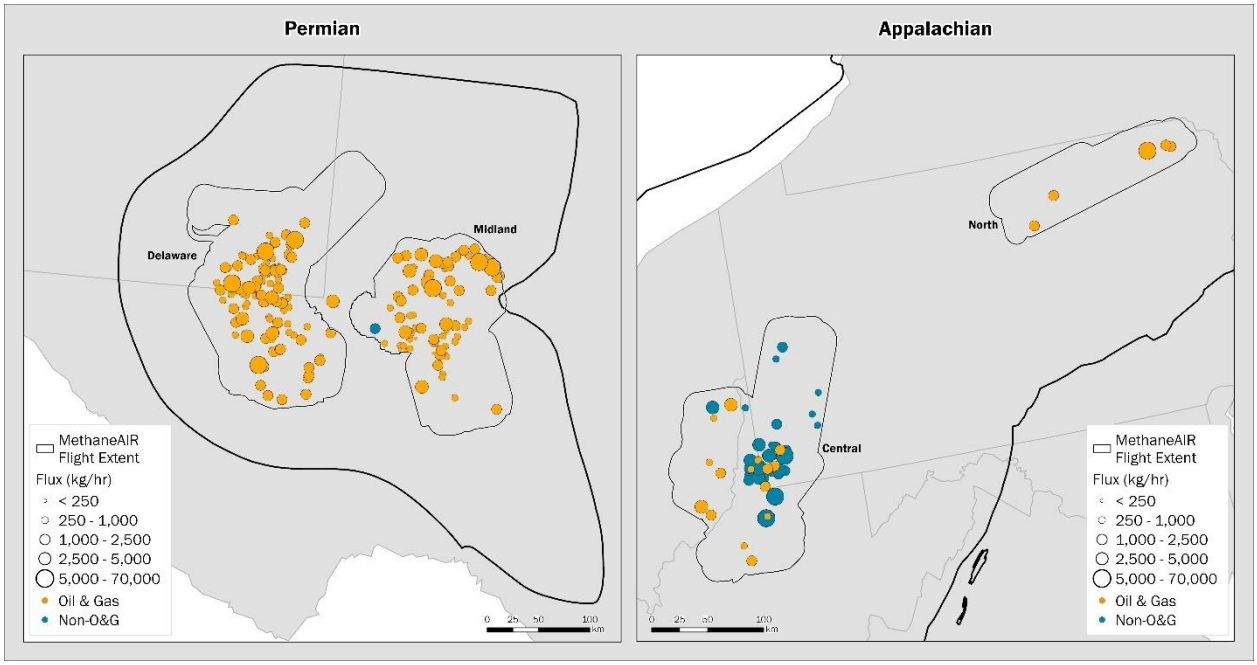


Figure S3.1 Discussed sub-regions within the Permian and Appalachian. Bold outlines indicate basins while gray outlines indicate regions flown by MethaneAIR without the inbound and outbound flight tails.

S4 . Calculation methods for point source emissions totals and MC Simulation

Studies using point source data to derive basin level estimates of emissions have commonly used some form of persistence, the number of times a source was detected vs observed to weight or average observed emissions (Cusworth et al. 2022; Chen Sherwin et al. 2022). The logic behind this weighting is that point source emissions magnitude can be highly variable both across its emitting duration and occurrence. Due to this variability, a sources contribution to an estimate of basin wide emissions should be weighted across all positive or negative detections or overflights. In the event where there are limited overflights of a given source, an additional step has been taken to use a monte-carlo simulation to randomly sample a distribution of persistence values according to observations from the same facility type and basin.

Our calculation approach for basin level point source emissions in this study is equivalent to a by-day persistence weighting approach, albeit without the additional step of monte-carlo simulations for sources with limited overflights. Our justification is that this study explores several basins either new or underrepresented the methane point source literature, and thus we do not have a preexisting extensive sample of representative persistence values to sample. Additionally, operator practices and emissions can change quickly over time, so it would be an additional assumption to apply persistence values from prior studies in overlapping regions.

The most heavily studied O&G basin in the literature is the Permian basin in Texas and New Mexico (Cusworth et al. 2021; Cusworth et al. 2022; Chen Sherwin et al. 2022; Kunkel et al. 2022). To test the possible effects of MC simulation in MethaneAIR basin level point source estimates, we compared approaches with and without MC simulation for sources with three or less flyovers. We used all available Permian persistence values collected by the AVIRIS NG and GAO platforms from 2019-2021 (Cusworth et al. 2022) and simulated persistence using the average across 10,000 random draws for the same facility type and subbasin. We compared the simulations effect on both basin-level total point source emissions magnitude and the relative contribution of facility types for observations from MethaneAIR in 2023 (Table S2.1)

Table S4.1. Comparison of monte-carlo simulation treatments for persistence weighting. Values in parenthesis indicate simulated 95% confidence interval. * indicates significant differences between treatments defined by no overlap in the 95% confidence interval.

	Average point source emissions for all sources (t/h)	Percent contribution 95% confidence interval by facility type				
		Compressor station	Wells	Pipeline	Processing plant	Unknown
Delaware						
MC simulation	47 (31 – 64)	19 – 32	16 – 26	4.3 – 71	3.7 – 16	2.1 – 6.9
No MC simulation	60 (42 – 78)	21 – 37	17 – 32	4.1 – 56	3.6 – 14	3.3 – 9.2
Midland						
MC simulation	*23 (18 – 29)	17 – 48	30 – 57	3.2 – 19	4.2 – 18	0 – 1.9
No MC simulation	*54 (42 – 65)	17 – 39	35 – 65	3.4 – 19	3.2 – 13	0 – 2.74

With MC simulation the overall magnitude of Midland emissions is significantly smaller such that there is no overlap with the confidence interval with the no-MC treatment. This is due to the fact that the four 2023 Midland flights yielded mostly areas with two repeat overflights of each source, and thus the entire Midland sample underwent MC simulation. In the Delaware, the MC treatment has again a lower emissions magnitude, but there is overlap in the confidence interval between

treatments. This is likely because the core area of the Delaware and much of the southern portion was overflown three or more times, and thus the majority of the sample did not undergo MC simulation in both treatments. Despite the shifts in the magnitude, the relative proportion of emissions by facility type remains similar across both treatments, with overlaps in the confidence intervals across all types. From this exercise, we conclude that the overall magnitudes of emissions should not be compared using the differing methodologies, however, we expect results comparing the relative sectoral contributions of emissions to remain the same regardless of methodology.

S5. Observational frequency and threshold for comparison

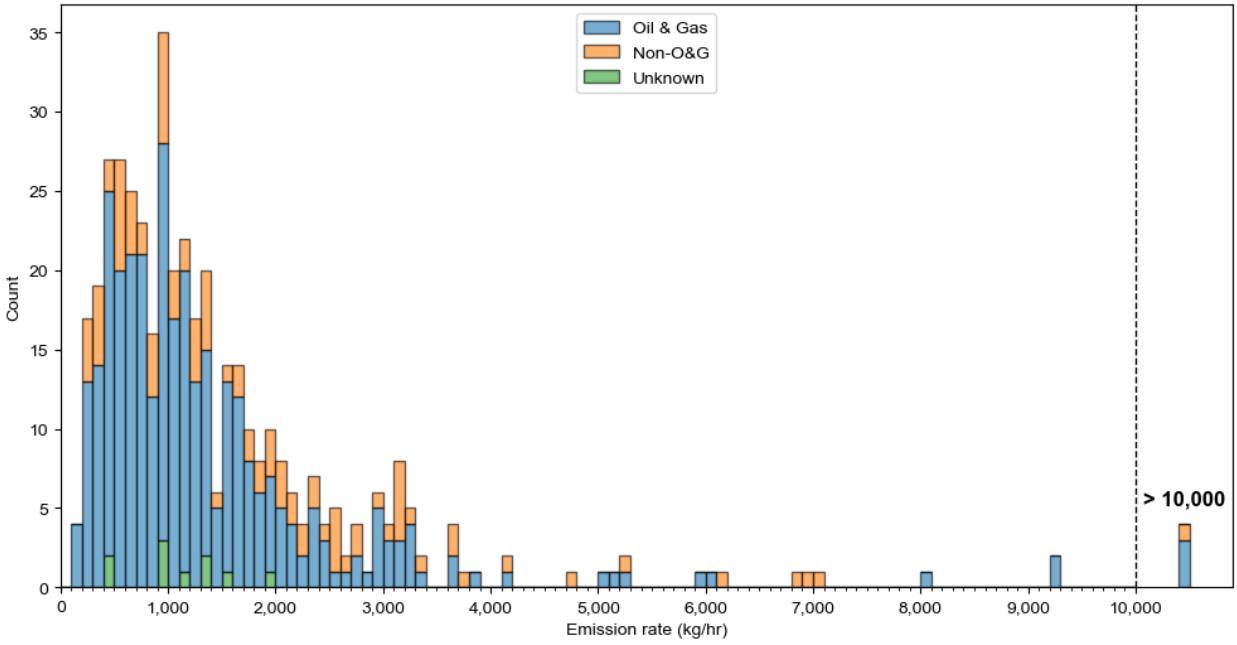


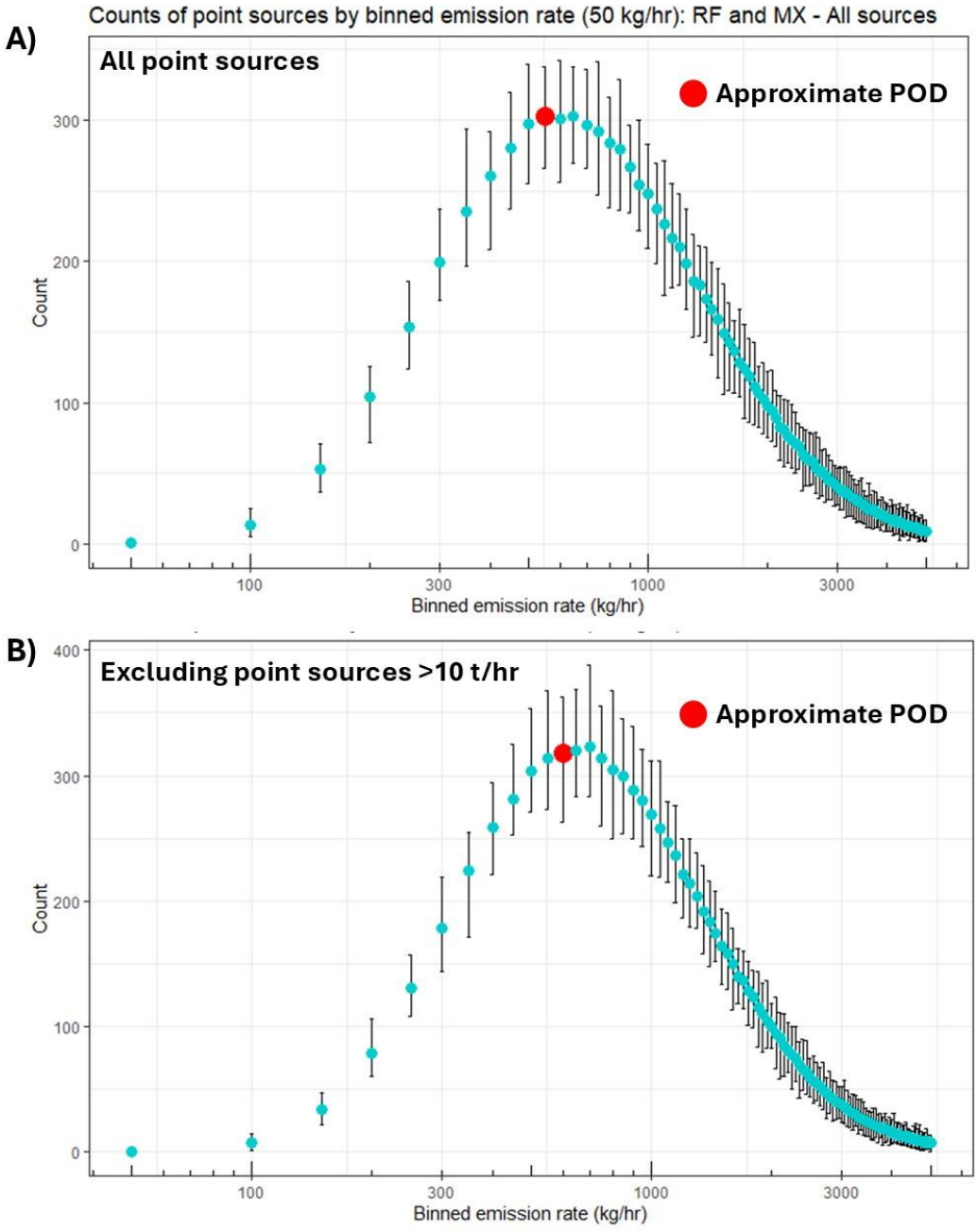
Figure S5.1 Observed emission rate distribution of all MethaneAIR detections by industry sector. Right hand side represents all plumes above 10,000 kg/h.

As mentioned previously, to meaningfully explore the relative contributions of sources, we must attempt to minimize the effect of present but unidentified sources across all scenes. Setting a threshold allows us to create a more definitive comparison that can be aligned with observations from other sensors if their sensitivity meets the same threshold. In single-blind controlled release testing, MethaneAIR has identified point sources as small as 33 kg/hr, but this detection likely did not represent a high probability of detection generally. Instruments do not have fixed probabilities of detections for point sources. Rather, the probability of detecting a point source is dependent on the emission source rate, in-situ conditions, and applied processing

methods, and thus variable with each emission source in a given scene. Determining the probability of detection for all point sources given the observed conditions in MethaneAIR campaigns, and if differing processing protocols using other instruments would perform at the same threshold is outside the scope of this work. However, we can take a simpler approach using some assumptions on the differences between observed and assumed emissions distributions. Anthropogenic methane emission sources follow a lognormal distribution (Zavala-Araiza et al., 2015), and thus the number of expected sources should increase as emission rate decreases. Therefore, the peak in observed emissions distribution should indicate a threshold at which unidentified sources begin to have an effect on observed distribution (Figure S4.1).

To estimate this emission rate threshold at which we observe a drop in the ability of MethaneAIR to confidently detect a point source emission, we used fit the empirical point source data from MethaneAIR to a lognormal distribution and extracted the disitrbution parameters. Next, we use the extracted parameters of the lognormal distribution to simulate 10,000 new point source emission rates, which were then sorted into 50 kg/hr bins from 50 to 5,000 kg/hr, which was repeated 1,000 times. We replicated this process again while excluding point sources emitting above 10,000 kg/hr to examine the impact of excluding these abnormally high emitting point sources relative to the majority of point source detections. We then assumed the emission rate bin at which the count of simulated point source emissions begins to drop as our estimated emission rate threshold.

Simulated emissions peaked at 550-600 kg/hr using all MethaneAIR point source observations, and at 600-650 kg/hr when sources greater than 10,000 kg/hr were excluded. Using the results of these two approaches, we set the threshold for comparison as a conservative value of 550 kg/hr since we observed no clear reason to account for the abnormally high emissions separate from the full suite of point source emissions. The probability of detecting a point source is dependent on several factors, such as the emission source rate, in-situ environmental conditions, topographical factors (e.g., land cover), and applied processing methods. Therefore, this approach for estimating threshold for comparison is likely a broad representation of the true emission rate threshold. Future work should investigate how the instrument's probability of detecting a point source changes across these variables.



240 Figure S5.2 Simulated distributions of point source emissions binned at 50 kg/hr intervals using A) all detected point sources and B) excluding point sources with emission rates >10 t/h.

S6. Temporal comparison of point source emissions across measurement platforms

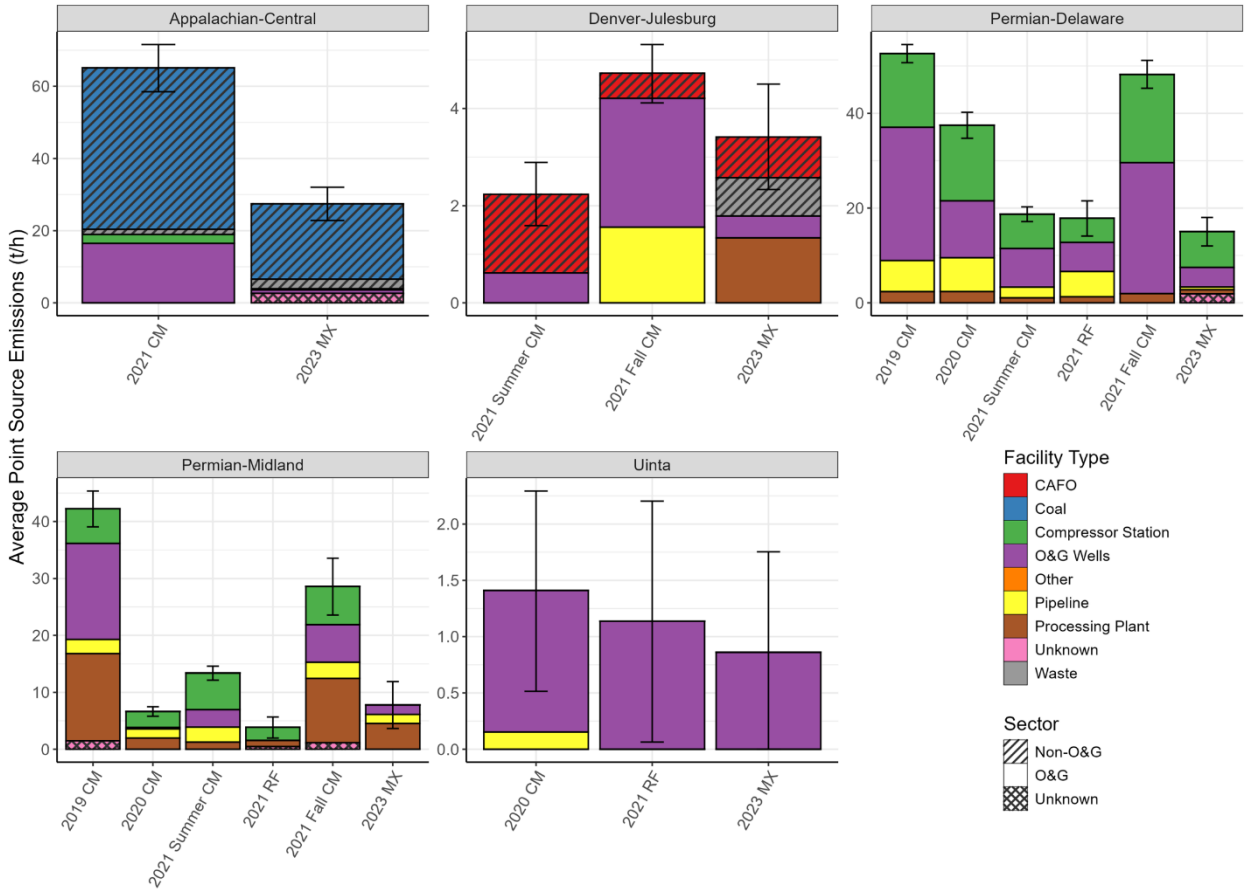


Figure S6.1 Temporal comparison of average point source emissions in shared study regions. Comparison is limited to observations above the threshold for comparison (550 kg/h) and from a shared sampling regions across all campaigns. Error bars represent simulated 95% confidence interval. X-axis refers to individual measurement platform campaign with the following abbreviations, CM: Carbon Mapper estimate using the AVIRIS NG and GAO instruments from Cusworth et al. 2022, RF: MethaneAIR research flights phase, MX: MethaneAIR 2023 observations.