We thank the reviewer for their insightful comments. Responses are in blue, while the original comments are in black italics.

RC2:

This manuscript describes the results from a large MethaneAIR campaign, an aerial methane measurement method, across several U.S. onshore production basins. The results from this paper focus on the distribution of high-emitting methane point sources across different types of facilities, basin-level trends in point source behavior, and understanding the relative contributions of methane point sources across different sectors. I believe this manuscript provides data that will advance our understanding of methane emissions and distributions, but would benefit from additional analysis. It is not immediately clear what the novel contribution of this work is, and additional analyses could be conducted to make this manuscript more compelling. Some suggestions for additional investigation would be:

Both reviewers commented to some degree that the scientific question and corresponding central findings of this study are not readily apparent in its current form. The scientific question of this study was articulated starting from line 68 as, “we investigate how high-emitting methane point sources are distributed across facility types, and if any patterns of emissions emerge across different regions with varying mix of oil and gas and non-oil and gas methane sources”.

Per the reviewers’ comments, this broad question does not sufficiently highlight presented results and enable a reader to evaluate whether the study was successful in answering its scientific question(s).

To better address this shortcoming, we have restated the scientific questions in in the final paragraph of the introduction (lines 70-74) as follows…

- *(i) What is the estimated distribution and contribution of high-emitting methane point emissions for basins in the U.S., and have these basins been explored in the prior literature?*
- *(ii) For basins previously explored in the literature, is there evidence that high-emitting methane point source emissions have changed over time? How does the relative sectoral contribution of high-emitting methane point sources vary basin to basin?*

We believe this framing better matches the analysis, results, and discussion, and that there are clear connections between the questions asked and figures and the discussion presented.

New learnings from these investigations includes methane point source data from multiple basins either currently not present in the methane point source literature—Anadarko, Haynesville, Arkoma, Bakken, Powder River, Greater Green River—and multiple basins that have been studied in the past but lack either recent or detailed analyses of point sources—Barnett, Eagle Ford. Other studies, namely Cusworth et al. 2022 and Sherwin et al. 2024, present results on point sources from multiple basins—Appalachian, Permian, Denver-Julesburg, Uinta, Barnett, San Juan—for years prior to 2023. This study builds upon prior multi-basin work through the temporal comparison of 2021-2023 MethaneAIR data to overlapping sampled regions to Cusworth et al. 2022. We consider our findings on the temporal trends of point emission sources through 2021-2023—decreasing overall emissions in the Permian and Appalachian-Central, level emissions in the Uinta, and consistent overall magnitude but fluctuations in contributions from individual facility types in the Denver-Julesburg—to be novel contributions to the scientific literature on high-emitting methane point source characteristics.

As noted by the reviewers, the study highlights several basin-specific insights—such as the consistent frequency of point sources in the Haynesville, or the lack of persistence among oil and gas sources in the Denver-Julesburg. However, as the overall scope of this study is on U.S. wide methane point sources, pertaining to surveyed oil and gas regions responsible for over ~80% of onshore production, our central findings are on point source emissions total per basin and the proportional contribution from oil and gas versus non-oil and gas sources (Table 1). The prevalence of non-oil and gas point source emissions totals relative to the emission rate threshold for comparison, and how they make up most observed emissions in several basins is also a key finding that has not been explored previously at this scale using methane point source imagers.

We have made several edits throughout the manuscript to accordingly clarify the research question, novel contributions, and central findings following these ideas.

- Further explore the contribution of emissions from other (non oil and gas) sectors. Are there other previous studies from these sectors that could be included for comparison?

Other non-oil and gas sectors that are grouped together but not explicitly named in the figures including power plants, a biogas facility, and a fertilizer plants. Comparisons of the power plant and fertilizer plant observations to prior studies have been added to the discussion.

"MethaneAIR observed emissions from several other facilities not commonly considered for their methane emissions, including power plants, a biogas storage facility, and a fertilizer plant. Plumes were detected at both a coal-fired power plant in the Eagle Ford and a natural gas-fired power plant in the Barnett. Emissions detected at the natural gas-fired power plant (6.8 ± 2.8 t h^{-1}) exceed estimated methane emissions rate of uncombusted natural gas from typical stack operations in prior measurement based work (8-135 kg h^{-1}) (Hajny et al., 2019), suggesting MethaneAIR's observation was a result of upset conditions or a separate fugitive source."

"MethaneAIR also observed emissions from a fertilizer plant in the Anadarko that was previously sampled using mobile surveys in 2016 (Zhou et al., 2019). Emissions quantified by MethaneAIR in 2023 (1.1 ± 0.4 t h^{-1}) greatly exceed the prior estimated fertilizer plant emission rate from mobile survey sampling across two days (213 ± 118 kg h^{-1})."

-Further exploration of the ultra-emitters (Line 213)

Ultra-emitters (> 10 t h^{-1}), and their disproportionate effect on basin-level emissions magnitudes and relative contributions of emissions by facility types are discussed in the first three paragraphs of section 3.2 Ultra-emission events and basin-level point source emission characteristics. They are also discussed specifically in lines 215-216, noting that four ultra-emitters make up ~20% of all observed point source methane flux. We also go into depth about one detection, comparing its reported emission rate to a state agency to the emission rate quantified by MethaneAIR.

-For regions with multiple overpasses, how do they day-to-day statistics compare? Earlier in the manuscript (Line 137), it is mentioned that with large enough sampling you could assume representative sampling across space and time (ergodic hypothesis). Do you find this to be true

in regions where there were multiple overpasses? Can anything be learned about minimum requirements for sampling, and the equivalency between spatial and temporal sampling?

We thank the reviewer for the suggested questions. The line referenced above pertained to the applicability of random sampling persistence values through a Monte-Carlo approach for weighting emissions in cases where there were few overlaps, and not about the minimum sampling requirements for developing a representative estimate of basin-level point source emissions. Section 2.1 briefly discusses sampling requirements and how that was accounted for in the planning phase for the aircraft campaigns.

In light of these questions, we have added an additional supporting information section that explores the day-to-day variation in select subregions with the highest number of overflights to expand on how regional emissions variation and intermittency affects the number of samples needed for consistent or representative emissions estimates.

The below figures show total emission estimates for a large subregion of the Appalachian-Central flown on five days and another subregion in the Denver-Julesburg flown on six days.

For the Appalachian-Central subregion, daily and aggregate emission estimates all overlap with one another given their respective 95% confidence intervals. This result is in part due to the large uncertainty ranges on the day-to-day estimates. However, any combination of two days would produce a total flux estimate without uncertainty that is within the confidence interval of the five-day estimate. We interpret these results to mean that for the emissions observed by MethaneAIR using these processing methods, one to two sampled days would produce a representative estimate of point source emissions for this subregion. Whether or not this interpretation applies to the entire basin would depend on changes in the observed facility composition and source intermittency going from the subregion to basin scale.

Conversely, results in the subregion of the Denver-Julesburg show a much larger degree of day-to-day variation. Observations from 10-09-2023 and 10-13-2023 are comparatively lower than the estimate from 06-22-2023, which had a single processing plant plume (~8 t/h) contributing most emissions on that day. Additionally, we cannot rule out the possibility of changes due to seasonal variation in the approximately three-month time period between the first and last flights. The range in the flux estimation without uncertainty using any combination of five out of six sampling days produces a larger range than that of the 95% confidence interval of the estimate using all six days. Assuming the observed variation is strictly due to the intermittency of methane point sources and not seasonality, we interpret this result to signal that several sampling days, possibly five or more, are required to estimate total point source emissions in this emission rate range for this subregion.

Our results in the Appalachian and Denver-Julesburg basins, although from different basins with very different facility type composition, suggest that required temporal sampling will vary by basin and possibly by subregion. This finding is consistent with results from Chen et al. 2024, which indicated that emissions characteristics in the production core vs periphery of the New Mexico Permian were significantly different, and that one comprehensive overflight may not be enough to estimate emissions depending on the desired accuracy. While only one daily estimate, 10-09-2023, produced an uncertainty range that did not overlap with the uncertainty range using all observed days, this result is primarily driven by large uncertainty ranges of daily and aggregate emissions estimates and not necessarily justification for the application of an ergodic hypothesis based on one comprehensive overflight.

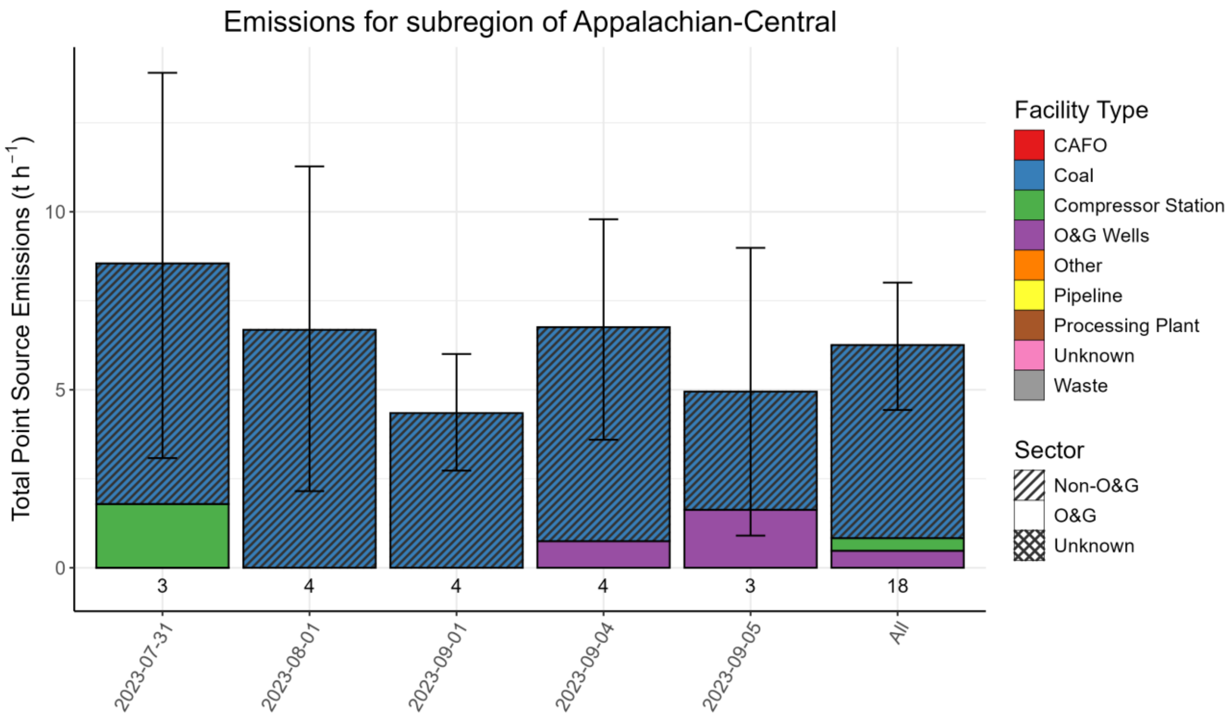

Figure S6.1 Daily variation of point source emission from a subregion of the Appalachian-Central. All represents the total point source emission estimate using all days with persistence weighting.

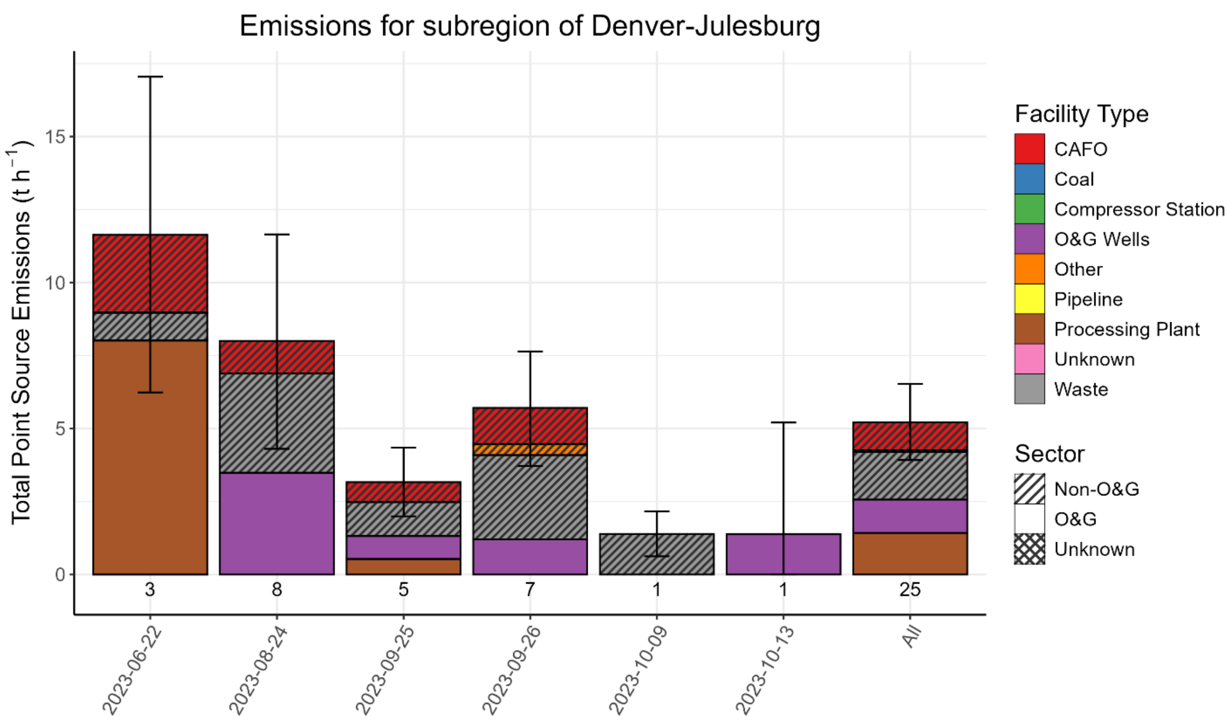

Figure S6.2 Daily variation of point source emission from a subregion of the Denver-Julesburg. All represents the total point source emission estimate using all days with persistence weighting.

In addition to the issue of scientific novelty, I also have specific comments related to the technical aspects of the work that I suggest be addressed before publication:

- Line 47: "Methane point sources above 10kg/hr can make up 13-59% of total regional flux" Please clarify the meaning of this statement. Are these point sources from all sources or just oil and gas sources? Is this just for oil and gas producing regions? What is the remaining 87-41% - is it from diffuse sources or source below 10kg/hr?

We thank the reviewer for the clarifying questions. The original statement, "Research by Cusworth et al. (2022) comparing point source quantifications to overlapping satellite-based inversions showed that methane point sources above 10 kg h^{-1} can make up 13-59% of total regional flux in the U.S" was a reiteration of the range of "Contribution of point sources to area flux (%)" values from Table 1 in Cusworth et al. 2022 alongside their stated upper end of their minimum detection limit in the study. These totals come from all sectors, including oil and gas, waste, coal, and agricultural emissions (CAFOs). In the study, the difference in flux between the satellite inversion results and the point source totals is the remaining 87-41% of flux, depending on the basin and individual flight campaign. We could interpret the 87-41% remainder as dispersed area emissions from sources both below 10 kg/hr and unidentified sources in their partial detection range. We have revised the sentence as follows:

"Research by Cusworth et al. (2022) comparing point source quantification to overlapping satellite-based inversions showed that observed methane point sources--above a minimum

detection limit of 10 kg/hr--from all sectors can contribute up to 13-59% of total regional flux in certain basins."

- The detection threshold is defined as 550 kg/hr, but data was included in the analyses and figures for emissions of ~200 kg/hr. Emissions below 150kg/hr were discarded. Please elaborate on this treatment of the data including the justification for excluding emissions below 150 kg/hr, while including emissions below 550 kg/hr. Is there detection testing that supports this?

To briefly summarize the thresholds as presented in the original manuscript.

Lines 120-121: "Plumes with a flux less than 150 kg h^{-1} were discarded as being below the detection threshold of the methodology. In addition, we manually reviewed all identified plumes and discarded plumes with known artifacts."

Using simulated XCH4 data, Chan Miller et al., 2024 found that the median detection limit for identifying plumes based on a MethaneAIR research flight scene (RF04) was 121 kg/hr (interquartile range: 106-141 kg/hr). This limit for identifying plumes is consistent with the 200 kg/hr threshold proposed by Chulakadabba et al., 2023, which was based on the results that estimated emissions below 200 kg/hr were uncertain and can be overestimated relative to known emission rates in a controlled release experiment. Based on these limits for identifying and quantifying plumes, we define the detection limit for this methodology at 150 kg/hr and discard quantified emissions below said threshold.

Mentions of the detection limit being ~200 kg/hr do not clearly reflect treatment of the data and have been standardized to 150 kg/hr in the text.

The emission rate threshold for comparison (Lines 159-162), defined by the peak observational frequency (Supporting Information S5) using a similar approach as Kunkel et al., 2022 and Chen et al., 2024, was approximately 550 kg/hr. We use this threshold as an approximation for a high probability of detection across all observing conditions. A standalone work is in development for quantitatively exploring probability of detection, as was done in Conrad et al., 2023 and Ayasse et al., 2024.

Emissions in the range of 150-550 kg/hr therefore represent the partial detection range for this methodology. To clarify the effect the partial detection range has on our estimates of total emissions, we have added some language in the main text and the below figure in the supporting information which is an alternate form of Figure 4a that shows the effective contribution from plumes above and below 550 kg/hr to basin level flux. Overall, the partial detection range has a minimal effect on estimated point source emissions, contributing only ~2.4% of the total flux estimate.

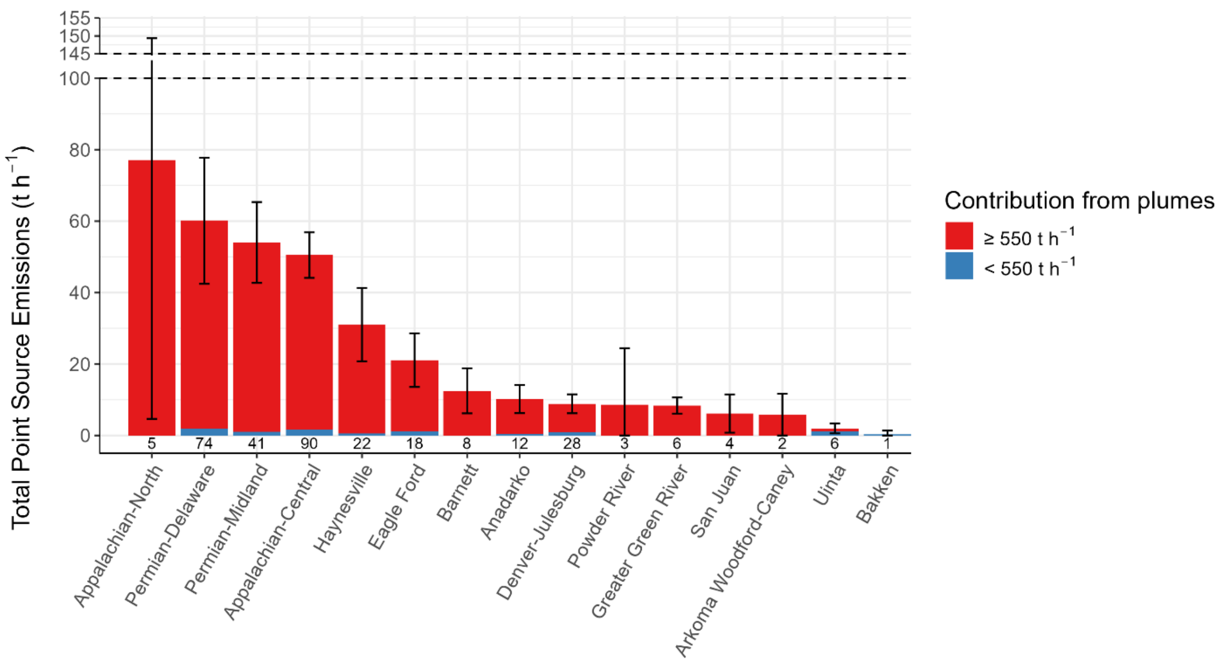

Figure S5.3 Average basin-level emissions totals for all high-emitting methane point sources detected by MethaneAIR in 2023. Colors indicate flux contribution from sources within the partial detection range (blue) and above the threshold for comparison (red). Numbers below the stacked bar chart represent total plume sample size. Error bars represent the 95% confidence interval of the basin-level total emissions for all high-emitting methane point sources. Note the axis break between 100 t h^{-1} and 145 t h^{-1}.

Additionally, language has been added throughout to ensure that resulting conclusions are recognized alongside the detection threshold for comparison, or in other words the portion of the emissions distribution curve that was observed.

Also in section 3.1, we have revised the previous approximation of detected emission rate range to actual emission rate range (160 kg/hr to 70 t/h)

I would also consider the probability of detection as well, rather than assuming a binary detect/non-detect. What is the probability of detecting a 550kg/hr source vs a 200 kg/hr source? If a site that was previously detected at 200kg/hr is not detected again, how can you discern whether this is because of a low probability of detection or because the source stopped emitting?

The assessment of a discreet probability of detection using a continuous function given source rate, wind speed, and flight altitude derived from controlled-release observations as done in Conrad et al. 2023, was outside the scope of this study.. However, research on estimating quantitative probability of detections for MethaneAIR using similar frameworks as in Conrad et al. 2023 and Ayasse et al. 2024 was presented at AGU 2024 (https://agu.confex.com/agu/agu24/meetingapp.cgi/Paper/1721700) , and is currently in development for a standalone study. Preliminary results indicate that 550 kg/hr represents a high (>90%) POD given the observing conditions of the 2023 MethaneAIR flights.

- Line 260: Related to above comment: Could the intermittency of detection be due also to probability of detection? Is it fair to conclude that just because the sources were not detected, they were not there? The manuscript states that "Recurrence of site-level emissions was seen only from non-oil and gas facilities" - was there 0 recurrence at oil and gas facilities?

The reviewer is correct in pointing out the possibility of emitting but undetected sources within the partial detection range and below the minimum detection limit. Lines 260-261 have been edited accordingly and additional language added clarifying this distinction where persistence and individual facility level emissions are discussed.

Line 260-261

"Recurrence of site-level emissions was seen only from non-oil and gas facilities, suggesting that across the six overlapping flights all detected oil and gas facility level emissions were intermittent in the Denver-Julesburg basin in Colorado."

Revised to

"Recurrence of site-level emissions was seen only from non-oil and gas facilities. While it is possible that there were undetected but present emissions below this study's effective detection limit, our results indicate that emissions from oil and gas facilities were all single occurrences in this size class for six overflights covering the core part of the basin."

- Line 133: Please explain why the emission rate was divided by the number of overpasses. It might be more helpful to report the average emission rate for detected emissions and the number of non-detects separately. The number of "0" emission measurements, especially when there were previous emissions there, can be just as important as the measured emission rates. This also relates to the first comment of understanding whether non-detects/"0" measurements are a function of probability of detection or source intermittency.

Emission rate is divided by the number of overpasses when calculating a basin's total point source emissions in order to normalize the contribution from subregions with more overflights relative to those with fewer. The interest of this study is to characterize the relative contribution of high-emitting point sources by facility type and sector across major U.S. oil and gas basins. We minimize the effect of true absence vs present but not detected emissions for basin-level estimates by limiting portions of the analysis by the threshold for comparison. Comparisons on the relative contribution of facility types (Figure 4b) and temporal comparisons (Figure 5) are limited by the threshold for comparison for these observations (550 kg/hr). Total emissions magnitudes by basin (Table 1 and Figure 4a) are not filtered by the threshold for comparison, but the effect is minimal (see prior comment, 2.4% of total weighted flux comes from detections below 550 kg/hr) and is broken down in the additional supplemental section.

Note, plumes are filtered prior to weighting by overflights (e.g. a plume detected at 1 t/h flown over five times contributes only 200 kg/hr of flux to basin level totals in Table 1 and Figure 4a & b. While the effective contribution of the plume is below 550 kg/hr, it is still present in Figures 4b and).

While contextualizing the analysis and results relative to the threshold for comparison limits the impact from sources with a low probability of detection, intermittency and flux variation at the regional scale still influences our results on basin-level estimates. Per a prior reviewer comment, this is explored in the newly added supporting information section no day-to-day variability.

- Table 1: Does the % oil and gas relate to the count of detections or the emissions contributions Aggregated total flux.

% oil and gas relates to the emissions contributions to aggregated total flux. Headers have now been modified to "% oil and gas flux" and "% Non-oil and gas flux".

- Line 208: Is there a reason why the DJ basin has the highest frac of non-oil and gas in Table 1, but the lowest emission rates?

The low median emission rate for the Denver-Julesburg (Figure 3a) and high fraction of non-O&G gas emissions (Table 1) is driven by the fact that most of the plumes are coming from waste and CAFO facilities, which as shown in Figure 2a have the lowest median emission rates of all explicity named facility types. Proportion of total point source flux from non-oil and gas sources is called out in Table 1, while individual plume emission rates and oil and gas versus non-oil and gas attribution are visualized in Figure 3a.

- Figures 2 and 3: Blue/green colors are hard to distinguish from one another

Thank you for the suggestion. The color map for figures 2 and 3 have been updated to match figure 4 and 5 for consistency throughout the manuscript.

- Figure 4: It would be helpful to include n values here – for example, my understanding is that the emissions from the compression station facilities in Appalachian north came from 1 detection (n=1). I think the sample size is important to communicate here.

Thank you for the suggestion. Figures 4, 5, S6.1 and S6.2 have all been updated to include the total number of plumes included.

-Line 255: How many measurements of pipelines were taken? Are there less large point sources because there was less observation, or are they just less prevalent in pipelines?

33/423 (from Figure 2a) of the detected plumes came from pipelines. A comparison of the detected pipeline emissions normalized by the length of pipeline observed, as done in Yu et al. 2022, is not included and we believe it more suitable for an in-depth study on exclusively pipeline emissions. From figure 4b, pipeline emissions make up 1-51% of total point source emissions from plumes ≥550 kg/hr, depending on the basin. As discussed in section 3.4, relative contribution of emissions may change with the size class of the emission rate. Nonetheless, these proportions are similar to values reported in Cusworth et al. 2022 (table 2.)

for the Permian (9-23%), Uinta (34%), and Denver-Julesburg (7-28%), despite the lower detection limit of the ANG and GAO instruments to MethaneAIR. From figure 5 (previous S6.1), we note that the total point source emissions from pipelines can vary significantly temporally, as evidenced by the lack of pipeline plumes in this size class in the Permian-Delaware from the 2021 Fall ANG and 2023 MX campaigns relative to prior campaigns in the region.

In-depth investigation on whether pipeline source are more or less prevalent in a certain emission rate size classes could be accomplished by comparing the emissions distributions curve per facility type from this study (Figure 2b) to one generated using instrument(s) with a lower detection limit (as is done in Williams et al. 2025, but this study does not delineate pipeline emissions).

- I would include more information on the comparison of these measurements with other measurement campaigns. I think this needs to be explored more, and to comment on whether the MethaneAIR distributions of high emission point sources agree with other data sources. I also think the information provided on Page 14 could be organized into a table or figure for more effective comparison.

Thank you for the comment. The cross-platform comparison was organized into a figure in the supplement (Fig S6.1). It has now been moved to the main body of the text as Figure 5.

Comparing the emission magnitudes in the overlapping sampled regions for plumes ≥550 kg/hr, Section 3.3, we see agreement in the emission estimates with some variation that could be due to yearly or seasonal variation and intermittency of underlying sources. From section 3.3, 2023 total point source emissions results from MethaneAIR in the Denver-Julesburg and Uinta are consistent with prior campaigns by Carbon Mapper. Results in the Appalachian-Central and Permian indicate generally decreasing emissions over time. Possible seasonal variability may mask temporal trends, as these campaigns were not all at the same time of the year.