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

RC1

This study describes the results from MethaneAir airborne campaigns across multiple basins in the U.S., with emphasis purely on attributed point sources above the MethaneAir detection limit. The study clearly presents the algorithms, the attribution, the results, and performs inter comparisons with other studies. I have a few comments that need to be addressed before I can recommend for publication.

1. *Although this study represents (to the authors' knowledge) the largest airborne methane survey performed over onshore oil&gas in the U.S., it wasn't entirely clear what new learning was gained from such effort.*

- (1a) *One possibility is in their section on repeat sampling in Haynesville, where they show that point source frequency is possibly stable.*

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 quantification and sectoral attribution 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 multibasin 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).

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

- (1b) *Another possibility is that non-oil&gas point sources make up an equal fraction of emissions as oil&gas point sources. That is very interesting, but the authors need to address detection limit issues if that's the central thrust of the paper. For example, certainly there are a lot of oil&gas point sources not detected by MethaneAir that were there (hence the need to filter results for comparison with Cusworth et al., 2022). For other sectors, if the general distribution of emissions is higher than the MethaneAir detection limit (e.g., underground coal vents), then those would be readily detected while some oil&gas sources wouldn't be detected, hence over-inflating the non-OG contribution to point sources. In any event, a presentation of this dataset itself is interesting, but some clarity on the general scientific questions this study is answering would be helpful to contextualize the results.*

The reviewer is correct in noting that non-oil and gas point sources make up the majority of observed total point source emissions, but only in certain basins (Appalachian-Central, Denver-Julesburg, Powder River, Barnett) and only when using this study's emissions threshold for comparison.

We respectfully disagree with the suggestion that the contribution from non-oil and gas sources, particularly coal in the Appalachian-Central, are overinflated in this study. In our analyses, we account for present but unidentified sources by using the emissions detection threshold for comparison, assessed based on the measurement data herein, which in turn ensures that our assessment of the relative sectoral contributions of high-emitting methane point sources are robust for sources emitting above this threshold. We believe we are very forthcoming about what portion of the emissions distribution curve is being analyzed, while highlighting that the relative contribution of facility types to total emissions will fluctuate across the broader emissions distribution curve. This contextualization relative to the method's emission rate threshold for comparison does not detract from our study, as nearly all aircraft or satellite imaging platforms focused on point source characterization share this caveat or limitation

2. Line 85. A more recent MethaneAir paper cites 33-38 ppb for MethaneAir precision in the Permian Basin and Arizona. Why is there a discrepancy? Is 17-20 ppb consistent with what you calculated in this study? It would be nice to know the background standard deviation in each of the regions you surveyed as it contextualized the distributions of detections in this manuscript.

We thank the reviewer for noting the differences between the two studies. Chan Miller et al., 2024 33-38 ppb precision is based on L2 retrievals at a 20m x 20m image. Meanwhile, Chulakadabba et al., 2023 reports 17-20 ppb standard deviation of gaussian denoised L3 images at 10m x 10m pixels for a separate set of flights. While there are inherent differences between what is estimated Chan Miller et al., 2024 and Chulakadabba et al., 2023, we do expect variation in the retrieval precision from flight to flight. For the flights included in this study, we found the pixel precision for L3 images to be in the range of 25-35 ppb depending on the flight conditions and light levels for the scene. This result has been added to section 3.1.

3. Lines 108-114. The DI and clumping technique you describe for plume detection uses many hard-coded values (e.g., 600-m pixel squares, 12 pixel clumps, etc.). Is this approach exactly the same as what was summarized in El Abbadi et al., 2024? In other words, how sensitive is the algorithm to these assumptions?

Those hard-coded numbers were all tested over a range of values, and the chosen values represent what found the most true plumes with the least number of false positives in our testing. The algorithm is somewhat sensitive to the values chosen in that other values may find slightly more true plumes, but with a lot of false positives, or may miss more plumes. The clumping technique for finding plumes was not used in El Abbadi et al. 2024, because in that case it was a controlled release experiment, and the location of the plumes was known. This method for finding plumes without prior knowledge of source locations has not previously been published.

4. Related to comment #2 - by selecting 600m windows, you are making a prior assumption on plume size which then could bias your detection algorithm. Put another way, if you assumed 100-m windows (i.e., smaller plume sizes), would this result in more plume detections? Have you tested this?

Yes, we have tested windows of various sizes, and 600m windows seem to best match the scales of plumes we are able to detect and quantify with this instrument. The value was chosen as our testing showed that it produced the most true plumes with the least false positives.

5. Line 261-264. This is conjecture. I don't think that conclusion is defensible without citation or actual operational data.

We agree and apologize for the original phrasing of these lines without contextualization or citation. The discussion point has been revised as follows to cite both the regulations and a recent study that uses process model simulations to explore their possible impacts on oil and gas emission reductions.

"Prior research using process level models of oil and gas emissions that takes into account the changes in the regulatory environment suggests that production normalized gas loss rates have

decreased in response to regulatory requirements starting in 2014 (Riddick et al., 2024). The observed lack of recurrent emissions ≥150-550 kg/ h^{-1} *from oil and gas facilities could possibly be attributable to this regulatory environment, which has included an empirically based oil and gas methane intensity verification program that was adopted in 2023."*

6. Line 353 and elsewhere. You cite the detection limit for MethaneAir to be ~200 kg/h repeatedly throughout the text, but then restrict most inter comparison to emissions above 550 kg/h. Other studies name this specifically with a probability of detection or an effective detection limit. You've shown in this study that effectively the detection limit for methane air is 550 kg/h, so it would be more accurate to update every mention of ~200 kg/h to "200-550 kg/h."

We thank the reviewer for the suggestion and noting the needed clarity for emission rates mentioned.

The minimum detection limit of this methodology was defined as 150 kg/hr in section 2.2, which we determined based on both the estimated limit for identifying plumes in Chan Miller et al., 2024 and the limit for reliably quantifying plumes from Chulakadabba et al., 2023. 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.

Mentions of the method 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.

In section 2.4 (Analyses of high-emitting methane point source emissions), we introduce a threshold for comparison that is based on the methodology used in Chen et al. 2024. Chen et al. refers to this minimum threshold as the "size limit for direct comparison", so we refer to it as the threshold for comparison and the range of 150-550 kg/hr as the partial detection range. Beyond this nomenclature, we've attempted to contextualize results and takeaways with specific mention of the emission rate class (e.g. "for point sources above 550 kg/hr…").

Mentions of the estimated threshold for comparison have been shifted from methods section 2.4 to results section 3.1. Additionally, we have added another figure to the supporting information (Figure S5.3) that clarifies the effective flux contribution per basin from sources in this partial detection range.