<u>Response to Referee #1</u>

Referee- The authors have re-focused the motivation away from national and regional emission rates and loss rates. Instead, they have expanded on a previous conclusion regarding the use of emission rate distributions to inform us on the ability of different technologies to locate emitting facilities.

This work is still focused on compiling literature emission rates (previously published in Omara 2022, 2023, 2024) which are categorized by facility type and production rate. These emission rate distributions are scaled nationally and by region using infrastructure inventories (Enverus, OGIM database) and validated against remote sensing surveys.

The manuscript is well written and the subject is of suitable content for EGUsphere. The subject is timely as there is active discussion regarding how mitigation funding can most effectively be used to reduce fugitive emissions from O&G. The figures are well designed and informative and have been improved from the first version.

While I do find the focus compelling, I still think the results fairly incremental. The majority of the data and analysis has been published previously (Omara 2022, 2023, 2024). The description of the probabilistic model is still confusing (general comment 1) and results are sometimes presented inappropriately, e.g. the bootstrapping approach provides confidence intervals about the probability of a facility emitting <LOD and not the probability itself.

Response-We thank the referee for their comments. We understand that the probabilistic model could still be explained more clearly and have made additional changes to the text in the methods, which include some additions to Figure 1 that also address the bootstrapping procedure also mentioned above. As responded in our previous round of response, we have clarified and addressed the referee's comments that this work is a significant novel work compared to previous Omara et al papers.

Additional changes-

L221: "…we first use bootstrapping with replacement (n=1,000) of our empirical measurement data to simulate the frequency of finding an individual facility emitting above the method LOD…"

L223: "The results of the bootstrapping procedure represent a normal probability distribution from which we estimate the frequency of finding an emitting facility (above the method LOD) with associated uncertainty bounds."

L233: "Similar to the process of determining the frequency of finding an emitting facility, we use the results of the bootstrapping to develop a normal probability distribution that classifies an emitting facility as either a top 5% or bottom 95% emitter."

L242: "Loss rates are used to calculate emission rates for the top five highest production bins of well sites, whereas we directly estimate methane emission rates for the well sites in the lowest production cohort (Figure 1), and for midstream facilities excluding VIIRS flare detections."

Revised Figure 1 with the estimated mean frequency of finding a facility emitting <LOD listed.

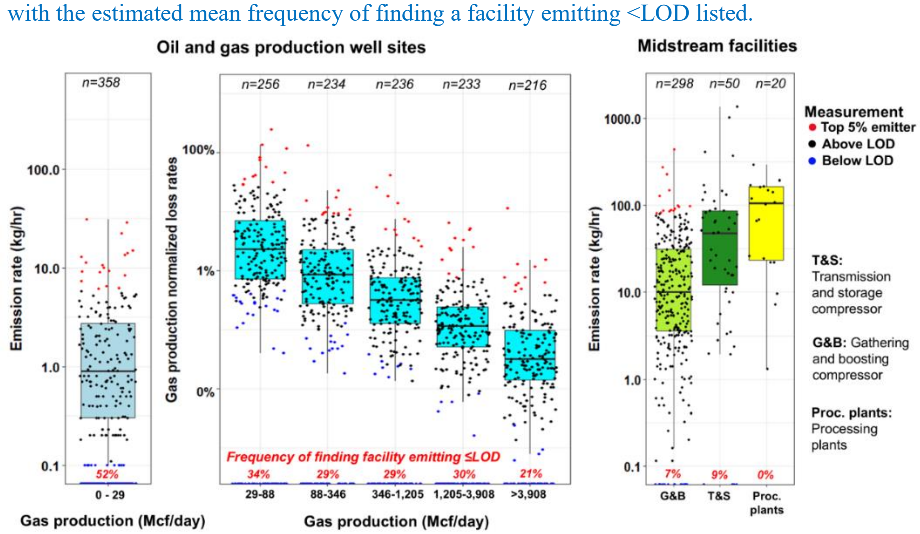


Figure 1 caption: "(...)The estimated mean frequency of finding a facility emitting below the method LOD is shown in inset red text at the bottom of each boxplot. We show absolute emission rates (kg/h) rather than normalized loss rates (%) for the lowest cohort of production well sites (...)"

Response to Referee #3

Referee- I require major revisions to the manuscript before I could recommend this manuscript for publication:
I recommend narrowing the scope of the study to focus on depth instead of breadth.
Don't try to bootstrap to basins, let alone the entire nation. This is unfounded and damages the credibility of the manuscript. Also, the most important findings in the paper can be supported without bootstrapping.
Instead, just analyze the aggregate distributions of the small sample size. To my knowledge, this is the first time the ground studies have been aggregated, and there are very valuable learnings from that alone.
Response- We agree with the referee that a more in-depth analysis of the empirical measurement data would be valuable and have added further details in Figure 1 that displays the estimated frequency of finding a facility emitting below the method LOD. We have further included direct analyses of the empirical measurement data in this work, as well as the sensitivity tests on the impacts of varying method detection limits and sample sizes as shown in Figures S8 and S9, and the general outlines and summary statistics of the empirical data in Tables S2 and S3.
We believe that a comprehensive assessment of facility-level emissions distributions must account for the diversity in facility-level characteristics that are inherent in each measurement data (e.g., for well sites: data collected over multiple years include a wide range in production rates and different facility-level emission profiles for various facility sub-types) and can vary from basin to basin. By incorporating oil and gas activity data at the basin and national level, our modeling approach, that takes into account of the compiled direct measurement data, makes it possible to effectively integrate these facility-level emission data over the population of methane emitting facilities, and draw more meaningful robust conclusions regarding the mean national- and basin-level distributions and potential implications for methane mitigation. Similar approaches have been used in the past (e.g., Alvarez et al,, 2018; Omara et al. 2018, 2022, 2024; Zavala-Araiza et al. 2015), in recognition of the limitation of simply consolidating empirical data from multiple studies without accounting for underlying factors that could impact each facility's emission rates, including the likelihood of a facility emitting below the detection limit or emitting at the "super-emitter" level.

Referee- Figure 1 and the related analysis are highly valuable if they can be trusted. I would like to see the authors provide more in-depth and transparent description of the data curation process. I can't tell by the descriptions of what was included and omitted whether there was cherry picking to support a foregone conclusion. I agree the with the authors' prioritization of studies that included zero emissions, but are there other ground-based studies that were considered and not included? Were all of the data from those studies used without omission? If all this checks out, then Figure 1 is very valuable. The only thing more I'd want to see from this data is what happens to the average loss rates in each category when the zero-emission measurements are included (which they should to give a true loss rate picture of the category).
Response- We agree that the data curation process could be explained more clearly and have added new text in the Methods that notes only one study that was not considered due to a combination age and being a component-level study (i.e., chance of "missing" emission sources). All other relevant studies were included to the best of our knowledge.

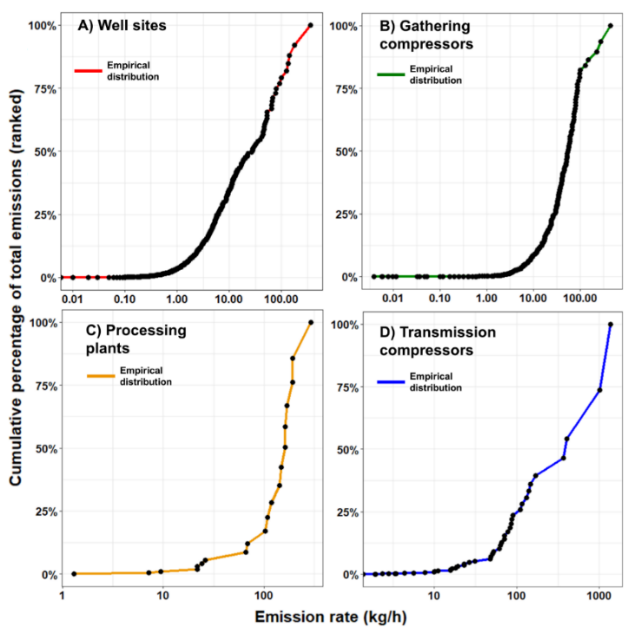
L123: "Only one study was excluded from our analysis (ERG 2011) due to a combination of age and a focus on component-level measurements."
We account for the influence of facilities with zero emissions in this analysis through the determination of the likely frequency of finding a facility emitting below the method LOD (i.e., the bootstrapping approach). If a facility is classified as emitting below the method LOD through our probabilistic modelling approach, it is then assigned an emission rate through resampling of all empirical measurement data below the method LOD (including zero measurements). Therefore, the loss rates from these BDL measurements are included in our analysis, meaning that the average loss rates would not change. For our results to be robust, it is important that we provide the best possible representation of what magnitude of emission rates would be encountered at individual oil/gas facilities, which requires that all relevant empirical measurement data are considered in our analysis.

Referee- Straight up analysis of the ground measurement cumulative emission distribution without any bootstrapping. Show the data so we can see how small the sample looks, especially if you want to break it down by basin and facility category. From that alone, what does that aggregate data tell us about emissions below <100 kg/hr. I can only assume the authors' thesis would remain unchanged: most of the emissions measured by the ground techniques come from emitters <100 kg/hr. If you want to use random resampling of the distributions to generate uncertainty, then that could work. The reach to say that finding is representative of a whole basin or nation is far too big.

Response- All the non-zero empirical emission rate data is shown in Figure 1 of the main manuscript, with the zero emission rates accounted for in the total number of emission rate data shown above each section of the box plots. We would also highlight that the dataset we use (~1,900 measurements) is the largest aggregated dataset of ground-based measurements from oil/gas facilities in the CONUS to date (to the best of our knowledge).

For information to referee, we include a figure below that shows the empirical distribution of the four main types of oil/gas facilities, which show agreement with our estimated distributions we present in the paper. For example, the percentage of cumulative emissions from well sites emitting <100 kg/h is 79% in our empirical measurement data, versus 90% (80-100%) in our estimated distributions. However, we note that the approach used in the paper is designed to account for the differences in individual facility characteristics using data such as the individual well site production data, for example, where site population and production rate characteristics vary within and among basins.

As we noted above, the resulting emissions distribution from a simple consolidation of all the empirical data, which are collected from multiple basins with varying distribution of facility types and emission profiles, does not comprehensively represent the population mean emission distributions, which is a more robust metric for understanding implications for methane mitigation. A "straight-up" analysis of the empirical emissions distributions overlooks these important nuances in the data, the effects of which are further likely compounded by our limited sample size, which we acknowledge in the manuscript. Our models and analyses account for these factors, and they provide a comprehensive picture of the mean facility-level emissions distributions.

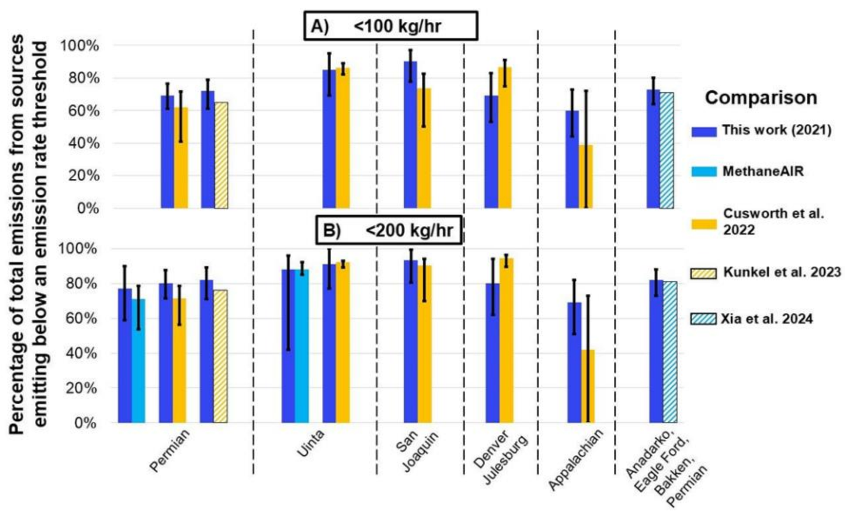


Referee- Rigorously and comprehensively compare the cumulative distribution from the small-ground-measurement sample size to those of the remote sensing distributions. The emissions scientific community needs such cross comparisons very badly, even with limited sample size and undetermined representativeness (this is all we have). I would like to see a detailed and transparent description of comparisons with Bridger, Carbon Mapper, and Methane

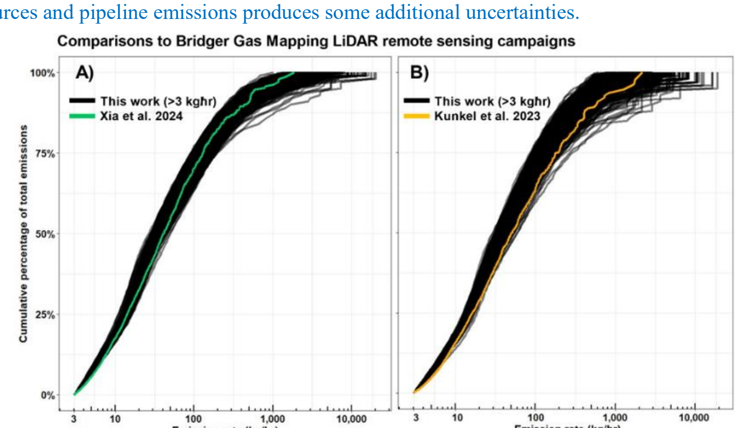
Air. I don't see how it could be possible that the ground distribution matches all three. That's okay, it will be one more piece to fill out the fuller picture.

Response- We agree that there is a need to perform these types of intercomparisons of aerial/satellite measurements to ground-based approaches – which is exactly what we have included in this paper, especially in the context of cumulative methane emission rate distributions, as this is the focus in the existing Section 3.4 and existing Figure 7 (see below) of our work. We would also highlight Figure S3 (see below) that shows a more complete intercomparison of the continuous emission rate distributions between this work and Bridger data from two studies, which is possible given the much lower 90% POD of Bridger compared to CarbonMapper or MethaneAIR. For our comparisons, we do indeed find good agreement on the distributions of emissions for all three aerial measurement platforms across multiple oil/gas basins, due in part to the steps we take to ensure a robust comparison of our results to those from aerial campaigns (i.e., restricting analysis to the same source types, ensuring the same spatial domains are used, focusing on emission rate thresholds that correspond to the measurement platform). We see the agreement between our results and the aerial campaigns as both validation of our probabilistic modelling approach using ground-based measurements, and that low-emitting methane sources in aggregate contribute a significant fraction of total oil/gas methane emissions. We have added additional text in the Methods that better describes this process, per referee's comment.

L273: "To perform these comparisons, we restrict our estimates and the results from other aerial/satellite studies to spatial domains of interest (e.g., an oil/gas basin boundary or the overflown domain from an aerial sampling campaign), and to specifically compare estimates of oil/gas methane emissions from the facility categories we are investigating in this work."



Existing Figure 7 – Shows strong agreement for all measurement platforms across different oil/basins, except for the Appalachian basin where the subtraction of non-oil/gas sources and pipeline emissions produces some additional uncertainties.



Existing Figure S3 – Comparison of continuous emission distribution curves to Bridger sampling campaigns which shows good agreement in emissions distributions above 3 kg/hr.