Author response to reviewer comments

Anonymous Reviewer #2

Review of Williams et al., Small emission sources disproportionately account for a large majority of total methane emissions from the US oil and gas sector

The paper examines the role of super emitters in the contribution of different midstream and upstream site categories in the oil and gas sector with regard to CH4 emissions. The authors use published data to create emissions distributions for the different categories as a function of daily gas production rate, using an algorithm that is very similar (identical?) to Omara et al. (2018).

The novel aspect in the present manuscript comes from the framing: Previous papers have stressed that super emitters dominate emissions in a given distribution of emitters, where the super emitters where generally the relatively highest emitters in a given distribution. The new manuscript emphasises the large role of emitters below a fixed absolute emission rate of <100 kg/h, so with this absolute definition, the super emitter category contributes less to total emissions that smaller emitters.

We thank the reviewer for the valuable comments and edits, and we hope the following responses address their concerns.

All of the page references in the responses below reference the attached manuscript with tracked changes included. Text in ***"bold blue italics"*** references prior text from the manuscript, and text in ***"bold red italics"*** references new added text.

My biggest concern is that this (and the title of the manuscript) imply a discrepancy between previous and new analyses, which is in fact not the case, and this is not well communicated. I strongly encourage the authors to clarify the change in perspective/framing and where it originates from. It is very clear from Fig 6 of the manuscript that in the category Well sites (<15 boed) there are almost no super emitters according to the absolute definition (>100 kg/h), so it is no surprise that they don't contribute much to emissions. And since this category contributes most to the annual emissions, (Fig 6b) the category strongly lowers the weight on super emitters to the national totals. This should be explained more clearly!

We note the reviewer's concerns on the perspective/framing of the paper and we agree that this is an important point to highlight and explain. Overall, the narrative surrounding oil/gas methane emissions has largely focused on the relationship between facility counts and total emissions, in the sense that a small percentage of the highest emitting facilities (i.e., super-emitters) contribute a disproportionate fraction of total emissions (which we also observe in this work). While these are important findings, our focus is on the relationship of emission rates to cumulative emissions, which we believe to be more critical information relating to methane measurement methodologies, especially with the rise of aerial remote sensing platforms with higher limits of detection. Furthermore, we present these results in the context of total national US oil/gas methane emissions, accounting for important information such as facility counts and facility types. As the reviewer correctly states, low-producing well sites are the major source of emission in our work, and individually do not emit at high rates, but cumulatively contribute a majority of total oil/gas methane emissions. Low-producing well sites also exhibit the same relationship of facility counts to cumulative emissions, since a small number of the highest emitting low-producing wells contribute the majority of their cumulative emissions. However, these "super-emitting" low producing well sites emit at roughly ~10 kg/hr, which is well below the detection thresholds of satellite point-source imagers, and also some aerial remote sensing platforms.

To better clarify the change in perspective/framing, and to better highlight the contribution of low-producing well sites, we have made the following changes to the main text.

- Text modified in the Abstract [page 1]
    - [page 1] ***" We estimate that production well sites were responsible for 70% of regional oil/gas methane emissions, from which we find the well sites that accounted for only 10% of national oil and gas production in 2021, disproportionately accounted for 77% (72-81%) of the total well site emissions."***

- Text modified in the Introduction
    - [page 2] ***"Several studies have recognized the importance of a small percentage of high-emitting sites (i.e. "super-emitters") and reported them as accounting for a large fraction of total methane emissions (Brandt et al., 2016; Cusworth et al., 2022; Duren et al., 2019; Sherwin et al., 2024). The emission rate thresholds that characterize these super-emitting facilities are critical information for methane measurement platforms, especially with the rise of remote sensing technologies that face limitations in detecting low-emitting facilities. Aerial and satellite remote sensing technologies have enabled more frequent monitoring of emissions from oil and gas sites and rapid mapping of large areas, although they do face limitations in detection sensitivity. "***

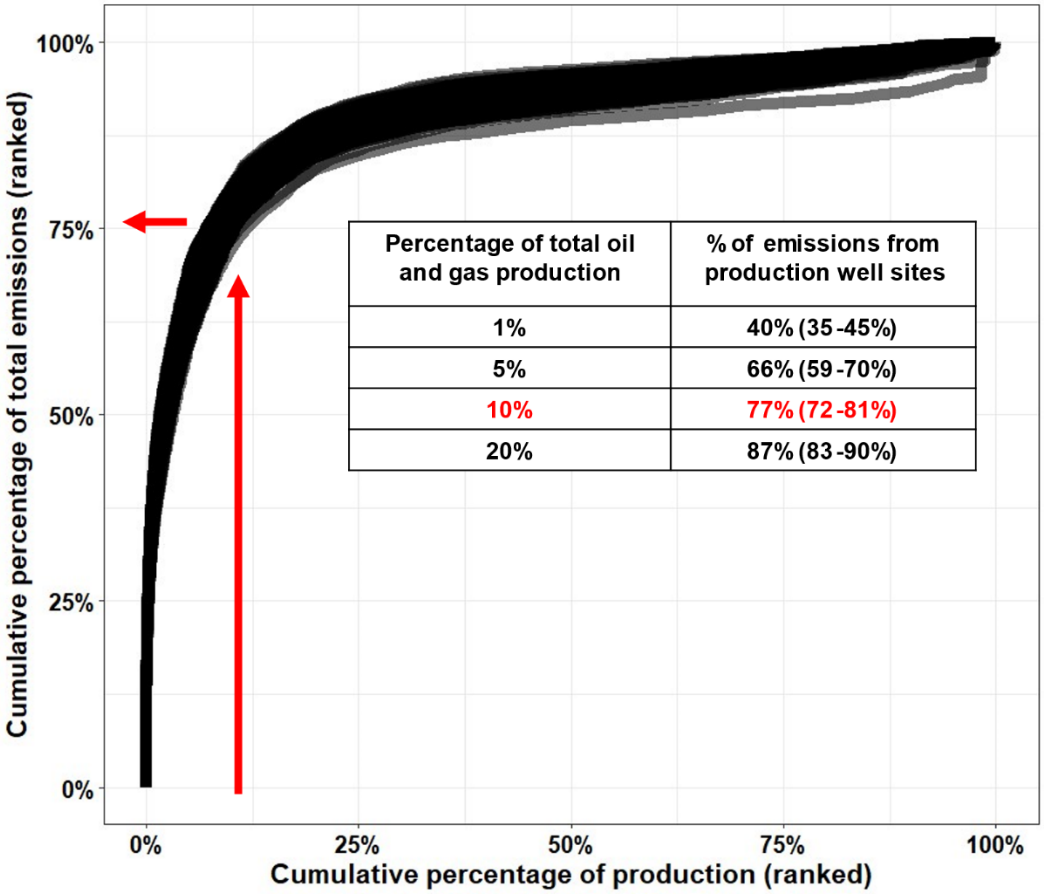
- Text added in Results
    - [page 20] ***"Production well sites constitute the bulk of total methane emissions among the facility categories we considered, with most of these emissions contributed from low production well sites. Overall, we find that 77% (72-81%) of well site emissions originated from only 10% of national oil and gas production in 2021 (Fig. S7), highlighting a disproportionately large fraction of emissions relative to production. In terms of individual well site level production values, the same 77% (72-81%) of total cumulative methane emissions were contributed from well sites producing 0.43 kt/yr (0.43-0.45 kt/yr) or lower. For well sites producing 15 boe/day (i.e., 0.13 kt/yr) or lower, which is the production threshold used to define a well site as being marginally producing in previous work (Deighton et al., 2020; Omara et al., 2022), we find that these low producing well sites accounted for 65% (58-69%) of total well site emissions, or 6.4 Tg/yr (4.7-6.8 Tg/yr)."***

- Text added in the Discussion
    - [page 25] ***"While detecting and mitigating emissions from super emitters are important (Cusworth et al., 2022; Duren et al., 2019; Sherwin et al., 2024), our results underscore the need to account for oil/gas methane sources emitting at lower rates, as the cumulative***

contribution of lower-emitting sites accounts for a large majority of emissions across US oil/gas basins. "

- Text modified in the Conclusions
    - [page 31] ***"3. Production well sites were found to be responsible for 70% of regional oil/gas methane emissions, from which the sites that accounted for only 10% of national oil and gas production in 2021, disproportionately accounted for 77% (72-81%) of the total well site emissions."***

- New Figure S7 relating cumulative well site level production to cumulative well site emissions



Percentage of total oil and gas production	% of emissions from production well sites
1%	40% (35-45%)
5%	66% (59-70%)
10%	77% (72-81%)
20%	87% (83-90%)

- ***"Figure S7: Results from 500 model simulations showing the cumulative methane emissions distribution curves for total well site oil/gas methane emission rates versus the percentage of cumulative combined oil and gas production. Results are ranked first by individual well-site emission rates, and then by well-site combined oil and gas production. The inset table shows the specific percentages of total emissions contributed from production well sites for cumulative well site production values of 1%, 5%, 10%, and 20%. The red arrows correspond to the percentage of total well site emissions contributed from well sites cumulatively producing 10% of total CONUS oil and gas production in 2021."***

My second recommendation is to clearly explain the underlying concept of the model. I read Omara (2018) again, and if I understand correctly, the approach is the following:

1) you use as input i) daily average gas production data for "all" sites and ii) a correlation between production data and measured emission intensity (percentage of production) from a small sub-set of sites, divided into categories.

2) you use a Monte Carlo technique to assign to each site in i) an emission intensity from ii) based on the emission intensity distributions in each category, to calculate hypothetic emissions per site, and then sum sites up in the different categories. You do that many times randomly to get statistically robust data.

In the present manuscript the essence of the model concept is buried in a lot of information on input data. Please state it more clearly.

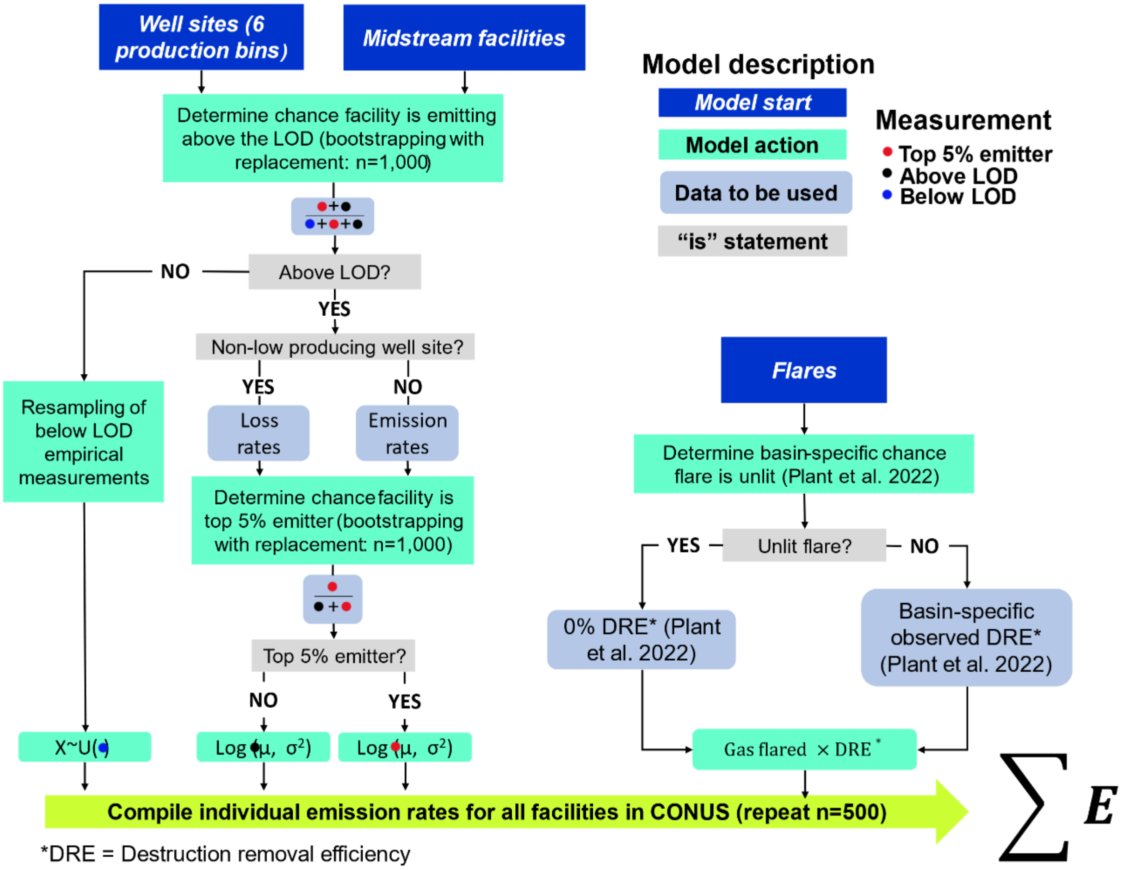
We acknowledge the lack of clarity and conciseness in the underlying concept of the model highlighted by the reviewer. Briefly, we use separate methods for 1) non-low production well sites, 2) low production well sites and midstream facilities, and 3) VIIRS flare detections. For non-low production well sites, we estimate gas production normalized loss rates for different production bins, which are then converted to methane emission rates using the annual-averaged individual well site gas production values. For midstream facilities and low-production well sites, we estimate the actual emission rates. For VIIRS flare detections, we use the published destruction removal efficiencies from Plant et al. 2022 and the gas flare volumes from VIIRS flares to calculate methane emission rates. In addition to changes to the Figure 2 of the main manuscript, we have also made the following changes in the text

- We have added new text in the first paragraph of the methods to provide a broader description of the essence of the model, and clarify that our methodology is based on multiple studies, and not just Omara et al. 2018.
    - [page 6-7] ***"We calculate annual methane emissions from all facility categories (i.e., six production bins of production well sites, T&S compressor stations, G&B compressor stations, and processing plants, and VIIRS flare detections) using a multi-step probabilistic modeling approach adapted from multiple studies (Omara et al., 2018, 2022; Plant et al., 2022) (Fig. 2). Briefly, for each individual facility and VIIRS flare detection in the CONUS for 2021, we estimate an annually averaged methane emission rate using empirical measurement data, and consequently the cumulative distribution of methane emission rates from the aggregation of these individual emission rates. Each emission rate estimate is indexed according to the corresponding replicate (n=500), and we use these repetitions to determine uncertainty for the cumulative methane emission distribution curves. The detailed steps of this process for all facility categories and VIIRS flare detections are described below."***

- New text added in the Methods concerning VIIRS flare detections
    - [pages 8-9] ***"For all VIIRS flares detections, we use the total reported volumes of gas flared for 2021 from flares detected using the VIIRS instrument (Elvidge et al. 2016) multiplied by the observed flare destruction efficiencies and percentage of unlit flares from Plant et al. (2022) to***

calculate annual methane emission rates from this source. As previously stated, our empirical measurements are largely located outside of oil/gas basins where the majority of VIIRS flare detections are located (i.e. Permian, Eagle Ford, and Bakken), but we cannot discount the possibility that there are instances of double-counting flares measured via our ground-based empirical data and those detected by VIIRS. For each VIIRS flare detection, we randomly determine whether it is an unlit or lit flare based on the basin-specific percentages of unlit flares reported by Plant et al. (2022). If a flare is determined to be lit, we use the corresponding basin-specific observed destruction removal efficiencies as reported by Plant et al. (2022) multiplied by the corresponding annual total volume of gas flared and convert to an emission rate. The basin-specific observed destruction removal efficiencies are estimated through a fitted normal distribution using the mean and standard deviations modeled from the 95% confidence intervals presented in Plant et al. (2022). If a flare is determined to be unlit, we use a destruction removal efficiency of 0%. For VIIRS flare detections located outside of the Bakken, Eagle Ford, and Permian basins, we used the total CONUS averaged flaring efficiencies destruction removal efficiencies of 95.2% (95% confidence interval: 94.3 – 95.9%) and percentage of unlit flares of 4.1% as reported by Plant et al. (2022).”

- Revised Figure 2 to simplify the decision chain of the calculations while also including the methodology for the VIIRS flare detections



*DRE = Destruction removal efficiency

- ***“Figure 2: Flowchart describing the facility-level estimates, with steps colored according to the specific process and data being used. We note that methane emission rates for flares are***

> ***calculated using a separate approach from that of production well sites and midstream facilities. Processing plants and T&S compressors are excluded from the determination of whether a facility is a top 5% emitter due to a lack of available empirical measurement data."***

Also, point 2 is making the method rather complicated, and I wonder whether one would not reach the same conclusions by deriving emission factors for the categories based on previous work, and using these emission factors per category for up-scaling.

The primary goal of our work is to characterize emission distributions across the full spectrum of methane emission rates, which requires an estimate of individual methane emission rates that are representative of what is encountered in the field. While we do provide estimates of total emissions at national/basin/aerial scales, in this case, a simple emission factor approach with upscaling by multiplying by number of facilities may produce a reasonable approximation of total methane emissions, but the individual emission rates by facility would not be representative of the actual methane emission rates.

We are attaching a response to anonymous reviewer #1 below, as we believe our response addresses these points:

- Multiplying an average methane emission factor by the number of facilities can produce a rough estimate of total methane emissions, but is not a suitable approach for characterizing facility-level methane emission distributions, which must account for the stochasticity in facility-level methane emissions profiles and related uncertainties (for references that discuss this stochasticity: https://www.nature.com/articles/ncomms14012, https://pubs.acs.org/doi/full/10.1021/acs.est.6b04303, https://pubs.acs.org/doi/full/10.1021/acs.est.8b03535) Developing robust methods for characterizing such distributions at the basin- and national-scale is the focus of this work. While we do present estimates of total emissions estimates at the national/basin/aerial spatial scale, these are a by-product of our methodology and not the main findings, which are the detailed distributions of individual facility-level emission rate and the large majority contribution of total emissions linked to an aggregate of smaller emitting sources (i.e., the distributions presented in Figures 3, 5, and 6).

  If, for example, an EPA GHGI emission factor (e.g., average methane emission rate per facility) and the associated confidence bounds (e.g., standard deviation of the mean) are applied to each individual facility to provide an independent emission rate, and this is repeated for all facilities in the CONUS, this simplified approach would not produce an accurate distribution of emission rates because a representative methane emission factor would still need to account for (i) facilities that may be non-emitting at any one time, (ii) the fact that different facility categories (including different production ranges of well sites) can emit at different rates at any one time, and (iii) the representativeness of facility-level empirical data (and inherent uncertainties in emissions quantification) when compared with the national population of facilities.

For these reasons, we believe a probabilistic modeling approach that accounts for these factors (and others) is essential to assessing emissions distributions and underpins the novel findings we present in this work. Moreover, the conclusions in terms of the specific emission rate thresholds and the aggregate emissions below those and their relative fractions to the total emissions across the US oil and gas upstream and midstream sectors as well as over each individual oil/gas basin has not been produced before based on empirically derived measurement-based analysis, which this study presents as a major step forward in our understanding the dynamics of oil/gas emissions and their source contributions which have important policy implications for measurement and mitigation, as we have discussed throughout the manuscript.

I strongly recommend using SI units, according to Copernicus guidelines (https://www.atmospheric-chemistry-and-physics.net/submission.html#math). I get confused by units like Mcfd, Mcf, boe, and cf3, in the text and in Eq. 1. I realize that these units are used in the O&G industry, but they should not be used in scientific publications. When common SI units are used, the rather trivial unit conversion factors can be omitted in Eq. 1, which would then read:

Emission rate = Gas production * methane content * loss rate * methane density

We agree with the proposed changes. However, we do think that it will be useful to retain the common oil/gas industry terms within the main paper for interested parties in the oil/gas methane science field who are accustomed to industry standards. We have included SI units and relevant conversions in both figure captions (when relevant) and in the main text, and also adjusted equation (1) as suggested by the reviewer. We hope these changes address the reviewers concerns.

- Changes made to Methods
    - [page 7] “***For the highest five gas production bins of producing well sites ranging from 29 to >3,908 Mcf/day (or 0.2 to >27 kt/yr of methane production per year, Figure 1), we use gross gas production normalized loss rates to model the distributions used to calculate methane emission rates from Eq. (1), where the: Loss rate is the fraction of emitted gas relative to gas production, the emission rate is rate of methane emitted from a facility in kilograms per hour, σCH4 is the methane content of the emitted gas which we assume to be 80%, and the gas production is the mass equivalent of natural gas produced in kilograms per hour at 1 atmosphere and 15.6 oC (1 Mcf = 1,000 cubic feet of natural gas = 19.2 kg of methane at 15.6 oC and 1 atmosphere; 1 boe = 1 barrel of oil equivalent = 6 Mcf). For the lowest well site gas production bin of 0 to 29 Mcf/day (i.e, 0 to 0.2 kt/yr of natural gas) and midstream facilities, we use the empirical absolute methane emission rate data as is. This approach is partly based on the methods used by Omara et al. (2022) for the non-low production well site category, which exploits a weak relationship between gross gas production data (which is most accessible in empirical measurement studies) and emission rates to better extrapolate emissions to the entire population of production well sites in the CONUS.”***

      Loss rate=(Emission rate [kg/hr])/(σ_(CH_4) × Gas production [kg/hr])   (1)

Except for these general points, I find the manuscript well written, but I have a few suggestions to the figures, partly linked to my general recommendations:

I find Fig. 2 complicated and it could be simplified in 2 aspects:

1) remove the "loop over i" going back to the start, and simply state that you do this for all facilities, (then also remove the index i).

2) show separate paths for categories 2, 1&3, and 4-9. I understand that this can be incorporated in an algorithm but this "high level" and rather trivial criterion makes the flow confusing.

These suggestions are very helpful, and we have since incorporated some of these changes into Figure 2 as suggested by the reviewer, including our revised approach for unlit and lit flares. The revised figure 2 was previously shown in a prior response to the reviewer.

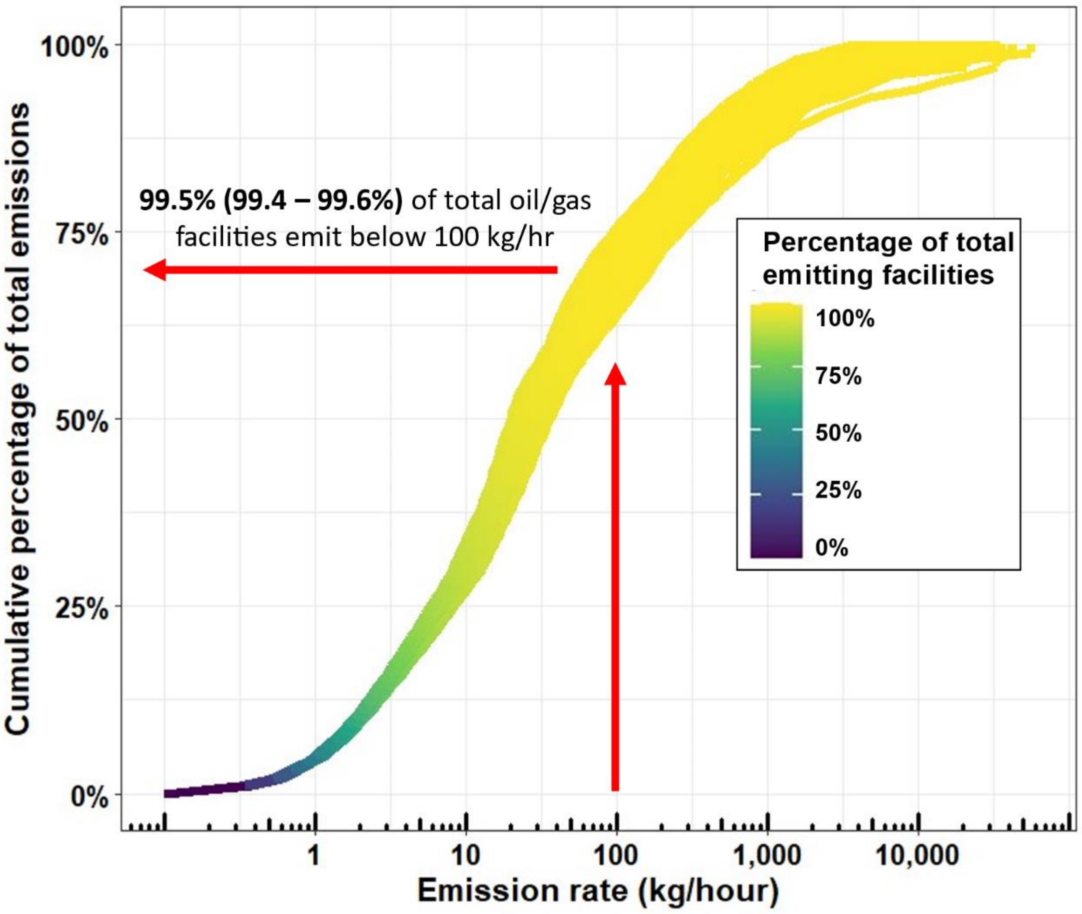
In figures 3,5 and 6 the cumulative percentage of emissions is plotted versus emission rate. In many previous studies the cumulative emissions were plotted versus fraction of total sites (ordered from high to low or low to high). These plots usually show the effect of the skewed distributions, namely that the highest emitters in a given distribution contribute most to the emissions. As I mentioned above, this apparent "discrepancy" should be explained, and it may help to also show the cumulative distributions versus fraction of total sites, at least for the category Well sites (<15 boed).

While it is true that many prior studies have shown relationships between the number of sites (ranked by emission rate) versus the cumulative percentage of total emissions, and that the top x% of highest emitting sites contribute a significant percentage of total emissions, our main goal is to define the emission rate threshold at which we define these "high-emitting" sites. For example, we can observe in a newly added Figure S13 that the same relationship between facility counts and cumulative emissions as highlighted in previous studies exists (i.e., small percentage of facilities contribute most to emissions). While this holds valuable information, it does not offer any information regarding the technologies needed to detect these "super-emitting" sites. We can determine that the top 1% of ranked emitting facilities contribute 37% of total emissions, and simultaneously that the bottom 99% of emitting facilities contribute 63% of total emissions. However, these analyses do not offer any information regarding the emission rate of these facilities, which is critical information when developing MRV policies. For example, the top 1% of emitting facilities (contributing ~40% of total emissions) have an average emission rate of 60 kg/hr. The fact that the top x% of facilities contribute a majority of emissions is a valuable insight. However, this offers no information regarding the emission rate detection thresholds necessary to measure/detect these facilities, which is the primary information we are communicating in this work.

In addition to the newly added Figure S13 in the SI, we have also made additional changes in the main text to better highlight this relationship (i.e., cumulative methane emissions versus facility counts or emission rates). We would also point to prior changes made regarding low-production well sites that we believe address the reviewers concerns.

- The following changes were made in the Discussion.
    - [page 25-26] ***"Most of our analysis centers around quantifying the percentage contributions of oil/gas methane sources emitting below one discrete emission rate threshold (i.e., <100 kg/hr, per EPA's definition of a super-emitter). We estimate***

that over 99% of the total oil/gas facilities that we analyze in this work are emitting <100 kg/hr (Fig. S13), which in turn contribute 70% (61 – 81%) of total methane emissions (Fig. 3). The emission rate threshold of 100 kg/hr is relevant to US policy decisions (EPA’s Final Rule for Oil and Natural Gas Operations Will Sharply Reduce Methane and Other Harmful Pollution., 2024), but we also illustrate a complete characterization of emissions, which gains importance as newer methane monitoring technologies have different LODs. For example, the effective LOD at high probabilities of detection for available point source imaging satellites of ~200 kg/hr (Jacob et al., 2022) would only be able to quantify 20% (10-32%) of all oil/gas point sources in the CONUS, if the full oil/gas sector was mapped in its entirety, based on our facility-level results. When considering the relationship of facility-level emission rates to total cumulative methane emissions, we find that oil/gas methane emissions in the CONUS are dominated by many low-emitting facilities, which relates directly to methane measurement technologies.”



○ ***“Figure S11: Results from 500 estimated facility-level emission distributions showing the cumulative percentages of total methane emissions contributed from facilities emitting below methane emission rate thresholds and colored according to the percentage of total emitting sites ranked by emission rate.”***

Technical points:

L127: …. intermittent sources…..

Changes made

L 185 and Eq. 1: It is not appropriate to use the chemical formula CH4¬ as the “methane composition”, in eq. 1. Refer to it as” methane content” of the gas and use a proper symbol

Equation has been corrected and simplified

L188: Omara (2020) is not in the reference list, should this be Omara (2018), otherwise add reference
Reference has been corrected

L534: reformulate

- Section has been rewritten, changes noted in a prior response