Author response to reviewer comments

Anonymous Referee # 1

Omara et al compiled previously reported methane measurement data to study methane emissions from major US oil and gas producing basins, and developed a high spatial resolution emission inventory for 2021. I find the methods solid and the manuscript well-written. I only have some minor comments.

We thank Reviewer 1 for these detailed comments and review of our manuscript. We provide point-by-point responses below.

General comments:

More information about spatial allocation from facility-level level to 10 km*10 km resolution is needed (section 2.4). As some basins have more data than others, how much uncertainty will spatial allocation introduce?

- Our methods for spatial allocation of mean total methane emissions requires information on emission estimates per facility and location of the methane emitting facility. Uncertainties on spatial allocation is therefore a combination of uncertainties in our mean total methane emissions per facility (which is discussed separately in Section 2.5 and in the Results and Discussion section) and uncertainties in the spatial accuracy and completeness of the methane emitting infrastructure location information. These latter sources of uncertainties are difficult to quantify based on available information. We also acknowledge that our spatial allocation represents the mean emissions estimates over the year [2021] and are not intended to characterize methane emissions at a specific point in time, as substantial temporal variability in emissions may be expected given the stochastic character of emissions.

- We have included the following sentences to provide additional clarification regarding our spatial allocation methods and related uncertainties [Section 2.4, page 9]:

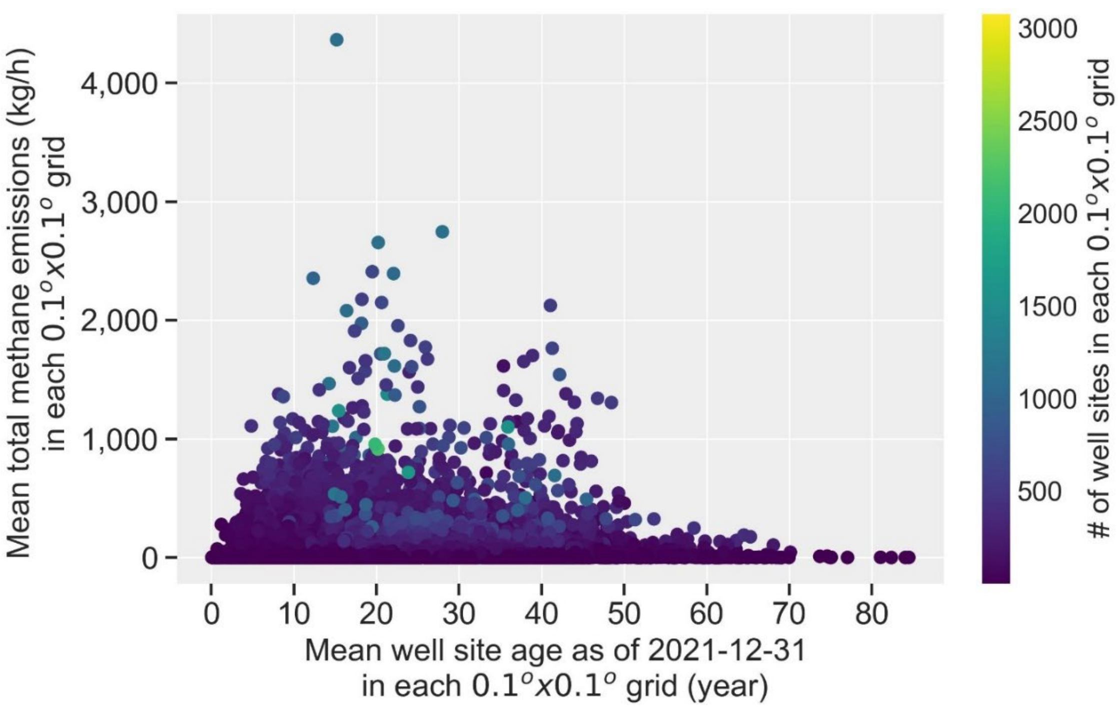
    "Our spatial allocation of estimated total oil and gas methane emissions is dependent, in part, on the completeness and spatial accuracy of oil and gas infrastructure locations for specific regions and oil and gas basins, for which related uncertainties are difficult to quantify based on available information. Our spatial allocation provides the mean methane emissions estimates for the year 2021 aggregated at each 0.1° × 0.1° grid (~10 km × 10 km) and are not intended to characterize methane emissions at a specific point in time, and are not intended to characterize methane emissions at a specific point in time, where substantial short-term variability in emissions may occur in part due to the stochastic character of facility-level methane emissions."

Additional analysis:

(1) As age of wells is important to methane emissions, would the authors add a plot to show the correlations between age of wells and methane emissions?

- The available measurement data on facility-level well site methane emissions do not generally include information on age of well sites at the time of measurement, and our model does not directly assess the influence of well site age on methane emissions. In general, the potential for fugitive methane emissions to occur at actively producing well site infrastructure is expected across well sites of varying age, a factor which contributes to a general lack of correlation of well site emissions with age (see, for example, Brantley et al., 2014).

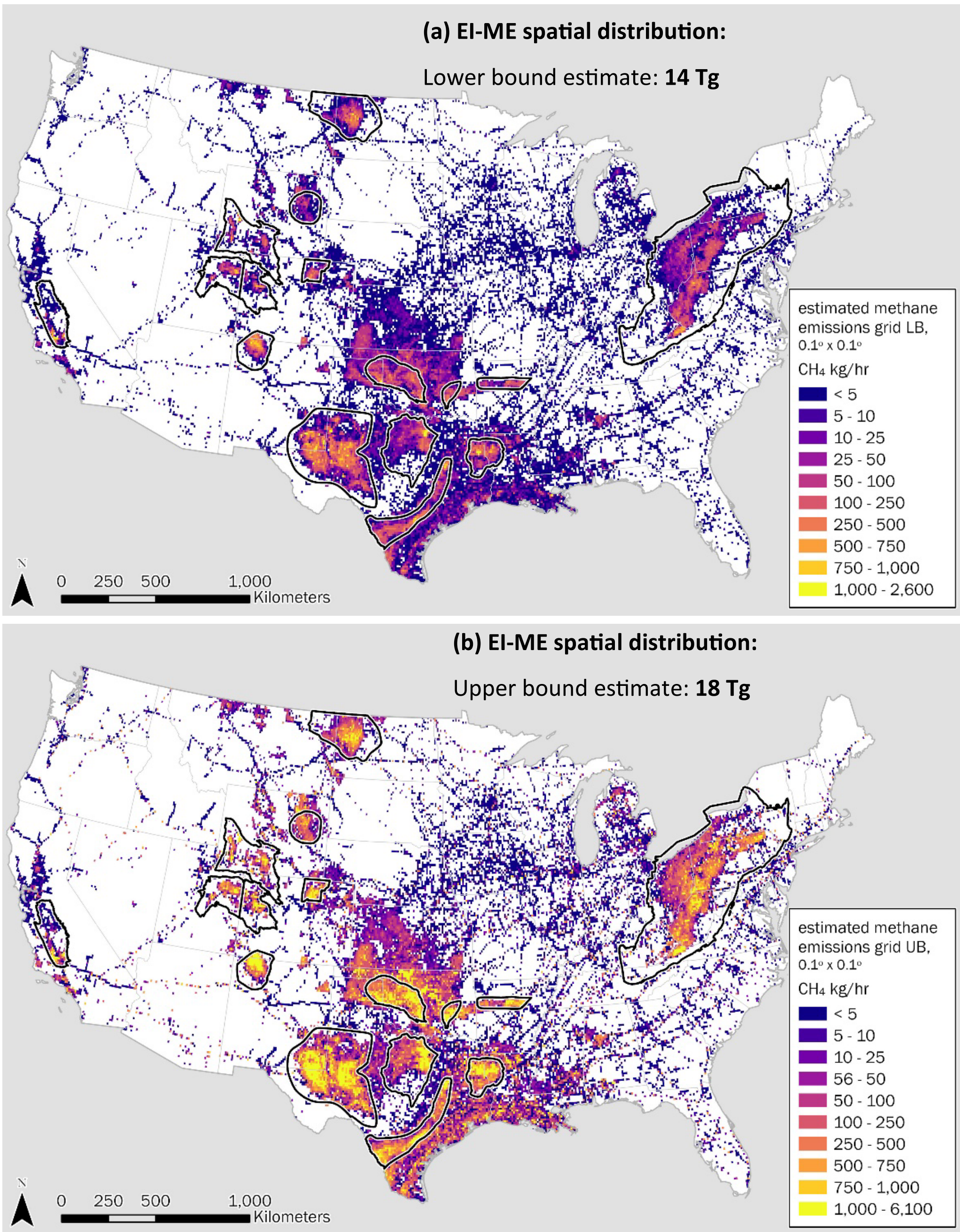
- Using proprietary data from Enverus Prism (www.enverus.com), we computed the mean age of well sites within each 0.1º × 0.1º grid (~10 km × 10 km) on which we aggregated mean total methane emissions. We compute the mean age as the average of the age of all actively producing well sites as of 2021-12-31. Figure AR1 below shows that there is essentially no correlation between the mean age and the mean total methane emissions within each grid cell, consistent with the findings from Brantley et al. (20214). Note, however, that we are not directly assessing correlations here because of limited information which precludes explicit treatment of well site age as a variable in our models (i.e., well site age is generally not reported in the measurement-based data).



Supplementary Fig. 8. Assessment of mean total methane emissions within each 0.1º × 0.1º grid (~10 km × 10 km) grid cell and correlation with mean well site age.

(2) It would be good to add a map in Figure 7 to show the uncertainty from EI-ME emissions.

We include in the Supplemental Information additional maps showing our lower and upper bounds on the mean total methane emissions within each grid cell.



Supplementary Fig. 9. Estimated spatial distribution of national methane emissions showing the confidence bounds on the mean total methane emissions: **(a)** lower bound estimate representing the 2.5th percentile within each 0.1º × 0.1º grid and **(b)** upper bound representing the 97.5th percentile within each 0.1º × 0.1º grid. The confidence bounds are based on 500 model simulations of each facility's methane emissions as described in the Main Text.

It's good to have a high spatial resolution emission inventory. Have the authors considered improving the temporal resolution of the inventory? If it is not possible, what are the challenges and how to make it possible?

- It is possible to produce the EI-ME inventory at finer spatial scales as our model simulates each facility's mean methane emission rates, which can then be spatially allocated to specific grid sizes if facility location is known. A higher-resolution version of the EI-ME inventory is available from the co-authors upon reasonable request for non-commercial, research purposes.

Specific comments:

Title: should specify what year is the inventory for.

- We have updated the title to indicate the inventory year:

  "Constructing a measurement-based spatially explicit inventory of US oil and gas methane emissions (2021)."

L21: should specify which year is the inventory for. And clarify what 'mean emission' represents (average of yearly emission, or average of all the uncertainty iterations).

- We have revised this sentence in the Abstract as follows:

  "We then integrate these emissions data with comprehensive spatial data on national oil and gas activity to estimate each facility's mean total methane emissions and uncertainties <u>for the year 2021</u>, from which we develop a mean estimate of <u>annual</u> national methane emissions, resolved at 0.1º × 0.1º spatial scales (~10 km × 10 km)."

L24: should add one decimal for '14-18' to be consistent with L23 '15.7 Tg'

- We have revised this sentence as follows, reporting our mean total methane emission estimates to 2 significant figures in Abstract:

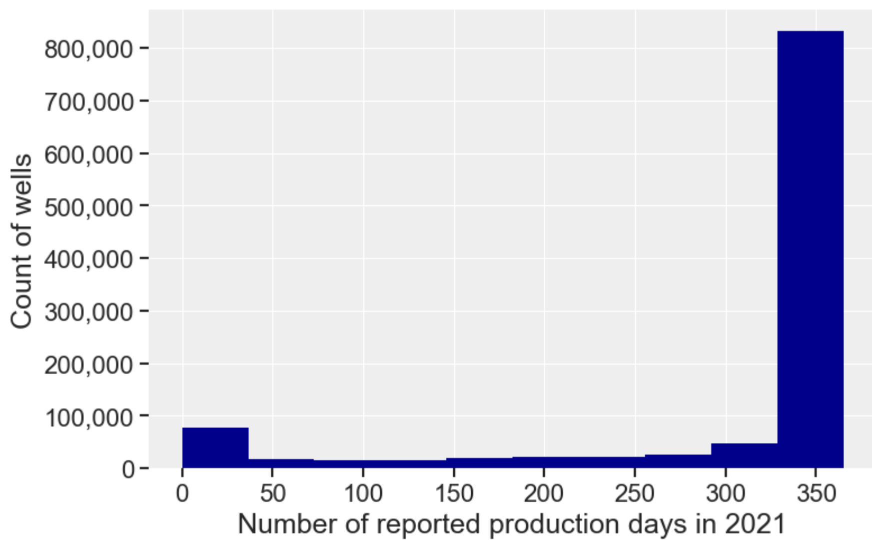
  "From this measurement-based methane emissions inventory (EI-ME), we estimate total US national oil/gas methane emissions of <u>approximately 16 Tg</u> (95% confidence interval of 14-18 Tg) in 2021 which is ~2 times greater than the EPA Greenhouse Gas Inventory."

L42,59: 'methane emissions' to 'methane emission', please check throughout the manuscript

- We use the plural "emissions" throughout the manuscript because the emission of methane arise from a variety of sources within oil and gas operations, for example, including from well sites to gathering and processing facilities to pipelines, each of which may have different root causes for the emissions (e.g., intentional venting, fugitive leakage, equipment malfunction, etc).

L84: how many wells do not have reported production days?

- Roughly 5% of wells in the database did not have reported production days, even as they reported production in the year. Figure AR3 below shows the histogram of reported production days per well, indicating that the vast majority of wells had reported production days.



Supplementary Fig. 10. Reported number of production days per actively producing well in 2021. Onshore US wells only. Analysis based on data from Enverus Prism (www.enverus.com).

- We have revised the following sentence in Methods to include the fraction of wells with no reported production days:

  "For each actively producing well, we derive average well-level oil (barrels per day, bpd), gas (1 thousand cubic feet per day, Mcfd), and combined oil and gas (barrels of oil equivalent per day; 1 boed = 6 Mcfd gas) production rates based on the reported number of production days, and assuming 365 calendar days in the year if production days were not reported, which occurred at <5% of producing wells (Supplementary Fig. 10)."

L104: need a little more information about how the previously published data are searched, such as keywords used for searching on google scholar (or somewhere else).

- Our focus was on previously published peer-reviewed data on facility-level methane emissions measurements for US oil and gas basins. Our search was conducted primarily based on Google Scholar, including key words reflecting the subject (oil and natural gas methane emissions), geography (US oil and gas basins), measurement methods (ground-based, OTM-33A, tracer flux), and major facility categories (well sites, compressor stations, processing plants, pipelines, crude oil refineries).

We have revised the following sentence in Methods as follows:

- *"We begin by performing a comprehensive data review and assessment of previously published peer-reviewed data on facility-level methane emissions measurements for US oil and gas basins, leveraging Google Scholar search results based on key words that reflect geography of interest (oil and natural gas methane emissions in the US), measurement methods (ground-based facility-level methods, OTM-33A, tracer flux, mobile transects), and major oil and natural gas facility categories (well sites, natural gas gathering and transmission compressor stations, processing facilities, pipelines, crude oil refineries)."*

L166-168: I'm confused with potential bias accounting, can you provide more information? And what about the uncertainty associated?

- One approach we use to evaluate the representativeness of well sites in the measurement data is based in part on the comparison of the distribution of well site gas production rates with the distribution of the natural gas production rates for the national population of sites. Overlaps in the two distributions provide confidence in the estimated results. As Supplementary Figure 3 shows, we find substantial overlap in the distribution for the sampled sites versus the national population, suggesting reasonable representativeness. However, the peak distribution for the sampled sites is greater than that for the national population of sites, suggesting potential oversampling of the higher producing sites in our sample. To account for this potential bias, we develop methane emission distributions based on production-normalized methane loss rates (methane emission normalized by methane production) with well sites stratified into seven different cohorts based on their production rates as described in greater detail in Section 2.3 and in Figure 1a. The uncertainties associated with our estimates is driven by uncertainties in the modeled distributions, which we assess separately for each cohort of sites as part of the model development. We provide further details for the uncertainty assessment in Section 2.5.

  We have moved this paragraph to Section 2.3 which describes the facility-level methane emissions model development and provided additional information on the comparison of the distribution of facility-level production rates for the measured sites with the distribution for the national population of well sites:

  > *"In addition, the distribution of facility-level natural gas production rates shows reasonable overlap with that for the national population of non-low production facilities, and the broad range in distribution of facility-level production rates across the national population of sites (~90 Mcfd to >50,000 Mcfd) is well represented in the sampled sites (Supplementary Fig. 3c). However, the distribution of production rates for the sampled sites suggests potential bias toward higher-producing sites relative to the national distribution (Supplementary Fig. 3c). We account for any such potential biases by developing emission models based on production-normalized methane loss rate*

distributions (methane emitted relative to methane produced) across seven cohorts of specific gas production rates (further details below).

We develop and use probabilistic emission rate distributions based on production-normalized methane loss rates, which shows a wide range <0.01% to >90% (Figure 1a) across all basins (Supplementary Fig. 3d), reflecting, in part, the diversity in production characteristics within and across basins. We use production-normalized methane loss rate distributions because (i) the empirical data across a wide diversity of oil and gas production facilities suggests an inverse relationship in which high-producing facilities exhibit comparatively lower methane loss rates, and vice versa (Figure 1a) and (ii) the consolidated dataset includes measurements collected in earlier years before 2021. By using the production-normalized methane loss rate distribution models for specific cohorts of facility-level production rates, we do not model any particular site that is active in 2021 as exhibiting the same emission rate size as observed when measurements were taken in the past, as the empirical data and the model constrains facility-level methane loss rates to production levels, which will be time-variant. As such, we provide a necessary constraint on our estimates, effectively adjusting modelled facility-level methane emission rates if production rates have substantially changed over time."

L221-225: As EPA inventory underestimates emissions, how does using EPA emission factors impact your results, and how is 50% uncertainty assumed?

- Our use of the EPA emission factors for pipelines makes it possible for us to estimate the emissions for these sources (for which we have spatial activity data) and provide a more complete inventory in the absence of detailed measurement-based data, even as we acknowledge that these emission factors are likely biased low. We assume a 50% uncertainty on these emission factors to be conservative, as EPA typically does not report uncertainty assessment for specific facility categories, while the reported uncertainty on the total emissions for all sources is generally approximately +/-20%.

- We have revised the following sentence as follows:

  "Given the scarcity of facility-level measurements for gathering and transmission pipelines, we use the emission factors estimated by the US EPA Greenhouse Gas Emission Inventory (EPA, 2022; 285 kg methane/mile/year and 582 kg methane/mile/year, respectively) and assume normal distributions of emission factors with 50% uncertainty. <u>Our use of EPA's GHGI emission factors for oil and gas pipelines makes it possible to provide a more complete spatially explicit inventory of oil and gas methane emissions (inclusive of gathering and transmission pipelines for which we have geospatial activity data), but likely increases uncertainties in our total methane estimates given potential underestimation in the GHGI emission factors."</u>

Figure 4: (1) why is the yellow bar for EDGAR hatched? (2) I suggest moving the year for each study from the bottom of the figure to the top.

We have revised Figure 4 as follows:

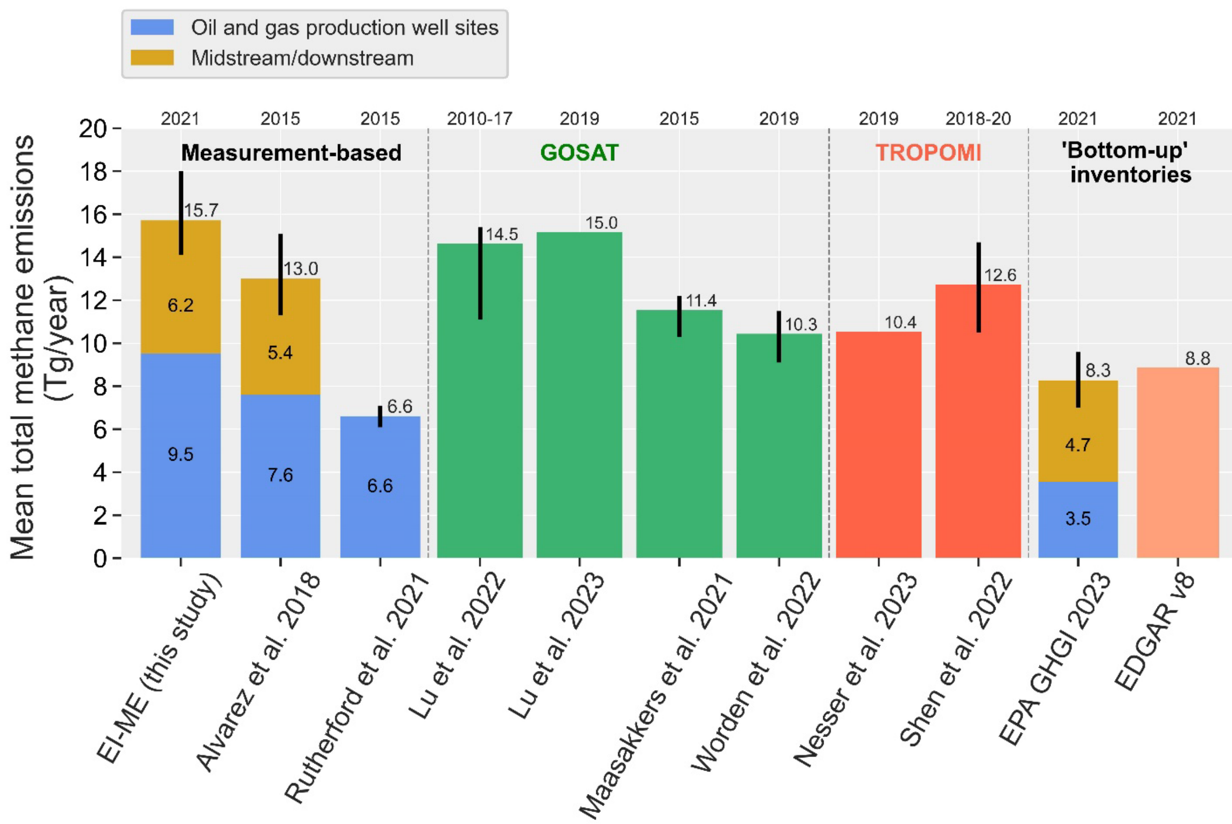


Figure 4. Comparison of this study's national estimate of total methane emissions from the oil and gas supply chain with previous measurement-based estimates and bottom-up inventories. The first three bars show the oil and gas methane emissions estimated based on facility-level measurements (this study, Alvarez et al. 2018) and production-sector-only methane emissions estimate by Rutherford et al. (2021) using models developed from component-level measurement data. Blue bars show the estimated emissions for the oil and gas production sector, gold bars show the estimated emissions for the midstream and downstream facilities (compressor stations, processing plants, refineries, gathering and transmission pipelines). Error bars show the estimated 95% confidence bounds on the mean total methane emissions estimates. This study's estimate of total national methane emissions include ~0.1 Tg/year of estimated methane emissions for Alaska. The green bars and the red bars show the satellite-derived estimates for contiguous US based on GOSAT and TROPOMI observations, respectively. The last two bars show the "bottom-up" inventories from EPA GHGI and EDGAR v8 for the contiguous US. In all cases, the year for which methane emissions are estimated are shown on the top x-axis.

Figure 5. The colors are similar and difficult to distinguish.

We have revised Figure 5, reducing the number of classes, for ease of readability.

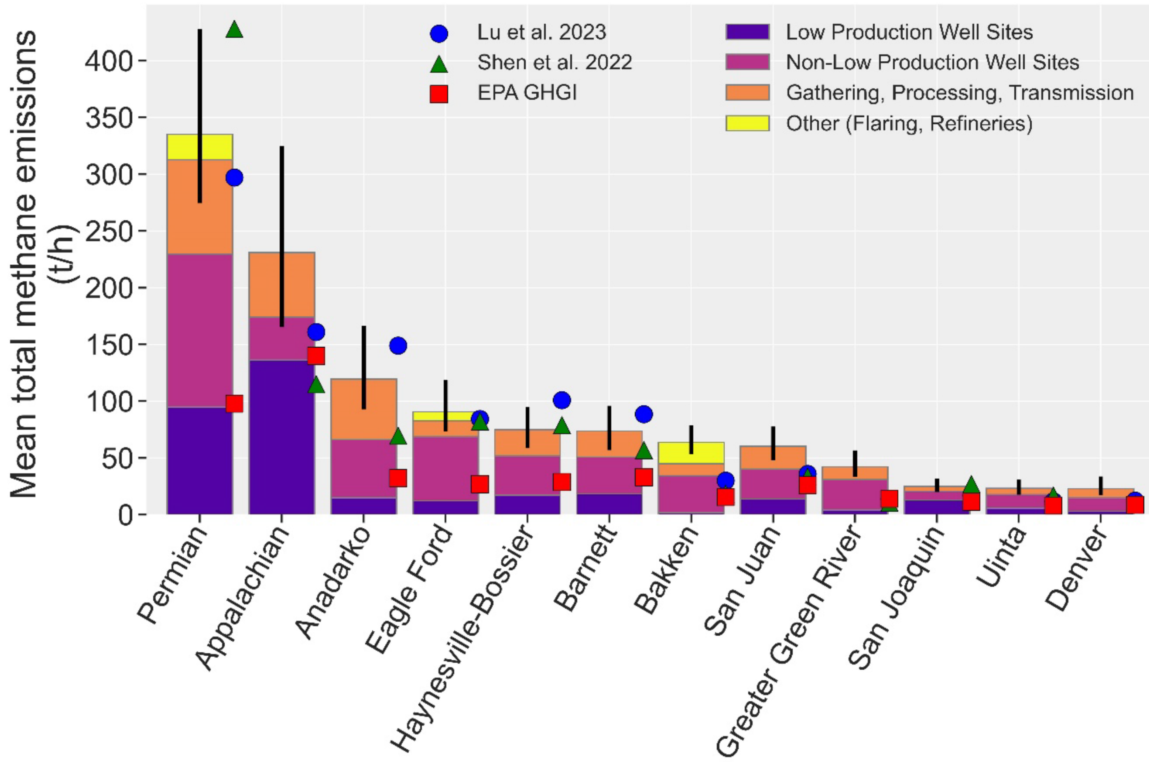


Figure 5. Basin-level differences in modeled mean total methane emissions and comparison with the EPA GHGI (Maasakkers et al., 2023), TROPOMI-derived estimates (Shen et al., 2022) and GOSAT-derived estimates (Lu et al., 2023).

L373: Miller et al.,2023 is missing in the reference list.

We have included Miller et al in the reference list.

Figure 7: consider moving the figure legends outside the maps (now they overlap with each other)

We have revised Figure 7 so that the legend is not intersecting the state/country boundaries.

References

Brantley, H. L., Thoma, E. D., Squier, W. C., Guven, B. B., Lyon, D. Assessment of Methane Emissions from Oil and Gas Production Pads Using Mobile Measurements. Environ. Sci. Technol., 48 (24), 14508–14515, https://doi.org/10.1021/es503070q, 2014.

EPA: United States Environmental Protection Agency, Inventory of US Greenhouse Gas Emissions and Sinks, https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks (last access: 20 December 2023), 2022.

EDGAR (Emissions Database for Global Atmospheric Research) Community GHG Database, a collaboration between the European Commission, Joint Research Centre (JRC), the International Energy Agency (IEA), and comprising IEA-EDGAR CO2, EDGAR CH4, EDGAR N2O, EDGAR F-GASES version 8.0, European Commission, JRC (Datasets), https://edgar.jrc.ec.europa.eu/report_2023, 2023.

Enverus Prism, www.enverus.com, last accessed February 06, 2024