Quantification of Oil and Gas Methane Emissions in the Delaware and Marcellus Basins Using a Network of Continuous Tower-Based Measurements

Response to reviewers

Review comments are **bolded.** Responses are in normal text

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

Overall the manuscript is very well written and gives thorough explanations of the model, the outputs and the errors, and provides important information currently beyond the abilities of satellite measurement. Quite often in the introductory sections I wrote comments only to remove them as I found the information somewhere later in the manuscript. It would be helpful to have some of this information earlier, such as the predominant winds and seasonal meteorological changes, and the inversion heights and diurnal cycles (with the associated SI figures coming much earlier) as this information is important when trying to understand the choice of tower location on the maps or the model input, rather than having to wait sometimes until the discussion for the information.

Otherwise my only other main point is that there are some SI figures that are not mentioned in the main text and others that do not appear in the order in which they are mentioned in the text, which makes for a lot of scrolling up and down when trying to consult both documents at the same time. Those figures and tables just mentioned in the SI text should be numbered after those in the main manuscript, as is Fig S21, but all should be mentioned somewhere in the text of either MS or SI.

Thank you for taking the time to review this manuscript. We hope the changes described in the detailed points below address your concerns regarding introducing helpful information when it would be most helpful, and not 7 paragraphs later. Additionally, we have revamped the supplemental figures, of which I am notoriously awful at organizing. For your amusement, here was the old order of the supp figures as they were called in the main text:

1, 12, 4, 10, 5, 2, 3, 7, 9, 6, 8, 13, 11, 22, 14, 15, 18, 16, 20, 21, 19

Not great. This has been fixed such that the new order goes from S1-23, and supp figures that were not called before are now referenced in the main text. Additionally, there are now only two figure (S17, 23) that are not used in the main text, but S17 is a Marcellus complementary to its Delaware S16 counterpart, so I've chosen to leave it there for organizational purposes despite not being mentioned in the main text.

Detailed Points / Questions

Line 59 – No explanation as to why a 4-hour period in the afternoon is taken. If it is taken as the period of maximum vertical mixing of the boundary layer does this work even in the winter months (at least 1 hour prior to sunset)?

Added clarity to this section, explaining that the 4 hour period is related to the time in which the boundary layer is at its maximum vertical mixing and reaches a stable height across all seasons, referencing the appropriate supplementary figure. We use information from the model abl heights to determine this time frame.

If daily afternoon averages are used for analysis, how can you have 70 and 40 background observations used per month as stated later. Needs to be clearer that the 4-hour window is just the starting point for background selection. Explanation comes too late.

Definitely some confusion here. The 70 and 40 are downwind observations (and are stated as downwind, not background). We have multiple towers in each basin that are downwind of the sources on any given day, so a day can have multiple downwind observations depending on the wind direction.

Line 64 – why did the Marcellus measurements stop at the end of 2016 as there is only 20 months to work with in the model?

Clarified in the text that funding ended. It's definitely painful to have an operating tower network capable of monitoring emissions in real time only to have to take it down because the project funding ends, especially in a place like the northeastern Marcellus which isn't a particularly great target for satellites to monitor due to its small signal and 140% chance of clouds.

Fig. 1 caption – using the phrase 3-km domain for the map is confusing. This is the model resolution rather than the area of the inset boxes, presumably those given in lines 70-71.

Agreed. Years of internally naming model domains by their resolution led to that. Removed 3 km everywhere in the text where resolution wasn't being referenced and just refer to it as the inner model domain.

Mention of the predominant wind directions would be useful given that it is stated earlier that the locations of the towers were chosen to be downwind of the O & G production areas, which one might presume meaning predominantly westerlies in the Delaware and easterlies in the Marcellus, but from looking at SI Figs 4 and 5 (not mentioned in the main text) this is clearly an incorrect assumption. Could refer to Fig S10 for the Delaware that highlights the seasonal differences.

First off, I do see in the opening paragraph of the tower description we wrote that the towers are designed to be downwind. Technically that's not quite right, as we design the networks to ideally have at least one tower upwind and downwind in most wind conditions. I've adjusted the sentence some to clarify that the goal is to surround the basin, rather than to establish all towers downwind of it.

In each paragraph describing the tower network, I have added a few sentences discussing wind direction and the primary role of each tower (with this role varying in the Delaware depending on season). Figures and supplemental figures are now referenced where appropriate.

Line 114 – It would be helpful to state the % of non-O&G sources in the Production-based inventory used in the Marcellus model. It must be much less than the 45% presumed from the PADEP, but if still 15-20% could errors in those emissions have implications for the model errors that should be mentioned in a discussion that is totally focussed on O&G?

Alright, this topic has gotten a bit of a makeover at the request of other reviewers as well. This changelog should address your concerns here as well. Last one is the big one, as we redid the inversion in the sensitivity analysis to solve for all sources instead of assuming non-O&G were perfect. The answer was unaffected.

Changelog:
-Added sentences detailing the non-O&G anthro sources in the Marcellus domain (it's mostly two landfills not co-located with the O&G).
-Added a figure in the supplement mapping out the non-O&G anthro sources. I did this for the Delaware as well to be thorough, although at ~1% other sources, that may be more for comedic effect.
-Added a couple sentences on the decision to neglect wetlands in our prior. The mean WetCHARTs ensemble has it at about 3 Mg/hr (O&G is probably somewhere around 20 Mg/hr in truth, and 7 Mg/hr in the PADEP inventory). There's admittedly no easy way to test for the influence this would have on our results given the coarse resolution (our study area fits only 6 grids of the WetCHARTs map) and huge uncertainty in wetland maps, but there's also no reason to suspect this 3 Mg/hr perfectly distributes itself in the production basin such that it biases the total in any meaningful way.
-Added a "Solving all sources" to the sensitivity analysis, where we allow the inversion to optimize non-O&G sources as well, and then extract the change to O&G by taking the total change in a grid and multiplying it by the fraction of O&G emissions that was represented in that grid from the prior. This scenario performed similarly to the default, with an emission total 8% smaller using the PADEP prior, and no change using the larger Production-based prior. This information has been added where appropriate in the manuscript, and the sensitivity analysis table has been expanded to include its results.

Line 204 – boundary layers not previously mentioned. Refer to Fig S8 which could have highlighted on it the period of the afternoon data that is used, and Fig S9, which could have highlighted the time of day in UTC at which the radiosonde is profiling the boundary layer.

Added references of figures to the manuscript. Added shaded area showing afternoon observational averaging period for each basin on the relevant figure. Captions in figure formerly known as S9 now references the time at which the observed and modeled comparison occurs (0 UTC).

Fig. 2 – the observed enhancements are influenced by the boundary layer height, but how does the modelled excess, based on the inventory account for this? Not clear, but presume that the 3000m summer and 1000m winter heights are used to calculate the difference in excess produced by the different Delaware inventories.

I'll admit I'm not entirely sure of the question being asked, but let me throw some comments your way and hope that one of them answers your question.

The observations obviously show the seasonality because, well… they are observations and are constrained by reality. I believe what you are asking is how the model accounts for the seasonality of enhancements based on boundary layer heights. That is addressed in the Lagrangian Particle Dispersion Model, which creates the surface influence functions. The dispersion model (at least the one used in our study) uses the turbulent kinetic energy (tke) fields from the transport model to determine how deeply the particles should mix vertically while moving backwards in time. On days with deep boundary layers (and thus deep tke fields), there are fewer particles at any moment that are within the "surface layer" (an upper bounds of a height which is defined as a surface interaction. I think it's 40m in LPDM but I'd have to pull out the innards of the code and check that number for certain). Fewer particles in the surface layer means you have a smaller surface influence function. Smaller influence function multiplied by an emission field produces a smaller model enhancement.

From here, the inversion is going to adjust the emissions to try and minimize the model and observed mismatch. The influence function itself accounts for that seasonality. Imagine a case where you have a 20 ppb difference between the obs and model both in summer and winter. In summer, when your boundary layer is higher, your influence function is a smaller magnitude, and the inversion will have to make a larger change to the emission field to make up that 20 ppb difference compared to winter where your influence function is likely larger, and thus only a small change may be necessary.

I hope that answered your question. It’s not really something that can be added to the paper itself, as it’s more related to how particle dispersion models function. I’m adding Lauvaux et al., (2008) to the manuscript references, which spends a full page talking about the particle dispersion model used in this study.

Now it’s also possible that I completely misinterpreted the question, in which case you are welcome to follow-up and ask it again, and I can take another shot at providing an answer.

Technical Corrections

Thank you for this subsection. I always find these after final submission and then stress out in a corner for 3 days. I’m sure there are other minor mistakes, but each one corrected before submission drastically reduces the number of days I spend stressed in that corner.

Line 115 – likely twice in 5 words.

Adjusted wording

Line 201-202 – open parentheses incorrectly positioned.

Grrrrrrrr LaTeX. Good catch.

Line 454 – basins twice.

Oh wow, I literally wrote basins basins.

Typographical errors on lines 13 and 132 of the SI text.

You told me what line to look on and it still took 3 readthroughs to spot the “the the”. Great catch.

Line 132 was a bit easier to spot. Fixed and fixed

<u>Response to Reviewer Referee 2</u>

The manuscript by Barkley et al. aims at quantification of methane emissions from two basins located in the U.S. (Delaware and Marcellus) using a combination of tower-based in-situ measurements, atmospheric transport modelling and prior emission data. The authors derive emission estimates and based on this, infer normalized loss rates for both basins. They find a temporal variability for emissions from the Delaware Basin, while emissions are found to be more or less constant over time for the Marcellus Basin. Overall, the study is well-designed and thorough. I really appreciate how careful the authors have conducted different sensitivity studies w.r.t. the impact of e.g. the background selection, the prior estimates and of intermittent emissions. I strongly recommend publication in ACP after addressing some minor points:

Thank you for taking the time to review the paper, and especially for acknowledging the sensitivity analysis! We really wish more studies would test the sensitivity of their inversions, as it could really help build confidence in the solution if the results are shown to be insensitive to the 72 subjective choices required to make them function. Below we've addressed your concerns and implemented the changes into the paper. Apologies for the 1 page background answer. I'm very passionate about that one.

Line 53: Please specify the type of the PICARRO instrument series

Added model information to manuscript.

Line 59: Please explain why you chose an 4h period and why in the afternoon. More details are found later in the manuscript, but I miss the explanation here.

Added a sentence explaining the reason when it's introduced (deepest and most steady vertical mixing leads to higher confidence that measured tower values are representative of full boundary layer column and thus less susceptible to model errors with mixing).

Line 95ff: What about emissions from other sectors? I guess they are negligible, but you should comment on this.

Emissions from other sectors are detailed just below this line. With that said, upon request the manuscript now goes into much more detail on other sources, particularly for the Marcellus where they're more than just 1% of the sources in the domain. Additionally, in the Marcellus the sensitivity analysis has been expanded to include a version of the inversion where we solve for all sources to see if this changes the solved emissions attributable to O&G. It does not thankfully (an expected result due to the lack of correlation between O&G and other sources in that region).

Line 136ff: I somehow struggle with the selection of the background. Thinking simply, I would choose measurements from the upwind tower. You mention that background approach later in the manuscript line (line 210f). I would expect a more detailed explanation of why you are not using this – straightforward - approach already here.

You struggle with the background because the background is a full page of wordage that I've written 8 times and still cannot find a way to easily express, made worse by the fact that using the simple "upwind" approach produces an almost identical answer, and the fact that you're the only reviewer to issue concern with it is frankly a miracle.

The issue is that the domain is so big and the winds so complex in the basin that none of the towers are really "upwind". Sure, if my winds are blowing to the east, I might look towards the tower to the west as "upwind", but the air takes ~10 hours to travel from the upwind tower to the downwind tower with our large network in purely steady state conditions, and the winds are going to veer in all sorts of directions during that period. This creates two issues: 1. The perceived upwind tower based on the winds might not actually be upwind at all, and 2. When looking at the upwind tower and downwind tower at the same hour, the upwind tower represents air with a history that's actually ~10+ hours removed from the downwind tower. You could theoretically use a lag on the upwind tower (i.e. use the value from 10 hours ago), but this leads to boundary layer shenanigans that have been found to have more of a detrimental impact on assessing the inflow than if you had just kept used upwind and downwind observations at the same time (Karion et al., 2021, Figure 4).

Taking a step back, the goal of the upwind in this study is to get a sense of the methane concentration of the air mass that enters our model domain. Once the air is in the model domain, the model resolves all the enhancements from sources within it and we solve for the emissions. In a perfect world if every source in our model was accurate, and our model transport was accurate, any tower could tell us what the inflow of methane was into the domain, because we could just take the observed methane measurement at the tower, subtract off all the enhancements from the model, and what would be left would be the methane mole fraction of the air that entered the model domain and reached that tower. Of course, we live in a very imperfect world, one full of errors both in the emission field and the model transport. The more enhancements from the model domain that must be subtracted off the observations to ascertain the inflow, the more likely we've added additional errors in trying to calculate that inflow.

Given that info, the easiest tower to calculate the inflow into the model domain is a tower where there are little to no enhancements from sources within the model. Most of the time, that is the upwind in the most traditional sense. Westerly wind? The tower on the west side probably has the smallest enhancement from sources in the domain. It's not quite always the case though. It's a big gas basin! You could have westerlies on one side of the basin and stagnant winds on the other, causing a large plume to accumulate on top of what would typically have been your upwind tower! That's no good, and that's what our derived background is trying to avoid. It looks at the 5 towers available and checks which one has the simplest inflow, i.e. has been influenced by the fewest number of sources within the model domain. And it goes with that…

…except, imperfect world. Sometimes the model will select a tower that's clean on the model grid, but in the actual observations records a massive 2500 ppb methane signal. Poor transport? Rogue cow? Regardless… that's probably not an adequate background choice. So a second check is performed; which tower recorded the lowest methane? The tower that recorded the lowest observed methane most likely experienced very little enhancement from sources within the model domain. That's it. That's the second check. But it too has flaws, in particular concerning the fact that inflow into the model domain is heterogenous, and you might just be picking a tower that's on the wrong side of a cold front for example.

Ultimately, each of these checks will work in some circumstances and fall flat in others, and so we take the average of both and call it a day. Well, not quite, as you saw we then do an analysis testing something like 6 other background methodologies just to be safe, including an extremely simple "use the wind direction" method. Taking the wind direction may be the safest and simplest method, and easiest to write about in a paper, but it also performed much worse in the model-obs statistics (final supplemental figure). Bayesian inversions hurt my head compared to looking at the obs-model difference and finding a scalar multiplier to apply to make them match, but ultimately I know the Bayesian inversion is going to perform better. The same is true for this background approach. I just wish there was a less dense way to discuss it. Honestly, the answer might just be to write a separate background paper for tower networks that cover large areas. I hope my advisor doesn't read that sentence.

For now, <u>I have added a simple explanation </u>to the intro of the background method on why just taking the wind direction is not a great method for our larger area tower networks. Addition highlighted in green.

"For example, the mole fraction value of a tower downwind of 0 methane sources within the model domain would be very representative of the methane characteristics of the air mass that entered the model domain and would serve as a good choice for a background value. Yet rarely do such conditions exist where tower measurements are not impacted at all by sources within the model. Towers upwind of the O\&G basin on any given afternoon can still have model enhancements overtop them, as winds are often not steady throughout the time air travels through the model domain, and plumes from sources that have traveled for hours or even days within the domain can sometimes exist overtop towers that would otherwise intuitively be thought of us "upwind". "

Line 463ff: This is an important, but very general statement. Please be more specific. Can you use your results to better design a network for different kind of emission landscapes? Or, if this is too ambitious (what I ´d assume), what kind of input could you deliver to someone who wants to set-up a network?

I like this comment! You're right that it's too ambitious, in that every basin will be unique in its meteo, seasonal winds, backgrounds, etc. BUT! All of this can be accounted for by running a transport model with methane tracers prior to setting up the tower network to get a better understanding of the size and structures of the various plumes that enter the region. Adding this to the paper.

Also a tangent on that note (you can skip): My first work was aircraft mass balance flights in northeast Pennsylvania. We ran a WRF-Chem run prior to flights there and anytime there was a southwest wind we'd see an enormous plume enter the study area coming from just a few coal mines about 500 km away, well beyond our area of interest, that completely overwhelmed the signal of the wells. We then learned that these mines were the largest point source emitters in the entire US and planned our flights accordingly to avoid those days. This is something we never would have caught had we not modeled the area in advance.

<u>Response to Reviewer Referee 3</u>

General comments

This manuscript presents results from a top-down inverse modelling study of regional methane emissions, using daily averaged mole fraction observations from two tall tower networks. Monthly fluxes are estimated for two oil and gas production regions in the USA, over two different time periods.

The manuscript's introduction clearly presents the motivation for the study, particularly in how this work could help form an independent estimation of emissions to aid with the monitoring of new emissions legislation. However, discussion of previous similar studies, and how this method and work differs from those, is lacking.

The authors find results comparable to those made by other top-down studies using aircraft observations over the same time periods and, similarly to previous studies, find strong discrepancies between their observed methane enhancements and modelled enhancements made using data from bottom-up fossil fuel emissions inventories.

The impact of intermittent methane sources on these regional methane fluxes is also considered. This is a novel development in this type of regional methane modelling and results from this test should be emphasised as a key conclusion, alongside the methane emissions estimates.

Overall, this is a clearly written manuscript, with a strong emphasis on detailed discussion of the study's results and sensitivity studies of key input parameters. However, some clarification on the method is required, particularly in how the observations are selected and used and in the assumptions made about non-fossil-fuel methane emissions.

This reviewer recommends that this manuscript is accepted for publication, after the minor comments below have been addressed.

First off, thank you for taking the time to review this manuscript, and we mean this as much more than just a formality. It's really hard to get reviewers this day, so we really do appreciate the time you dedicated for this. We agree with nearly every point raised in your comments and have made changes and additions to the manuscript to address them. We hope the comments below adequately address your concerns and make the manuscript ready for publication.

Specific comments

Line 41: An additional paragraph could be added to the introduction, highlighting how others have used similar methods for independent monitoring of regional methane emissions and how the work presented in this paper differs from these previous studies.

Added a paragraph detailing other aircraft, spectrometry, and satellite top down methods and their limitations, and expanded the final paragraph to give examples of tower-based networks utilized on urban scales for emissions monitoring and how this paper provides a similar concept but for O&G basins.

Line 59: Only daily afternoon averages are used from all observations, presumably due to issues with transport modelling at other times of the day? More detail on this choice of observation time and averaging is included in the supplementary, but the results from this sensitivity study should also be noted here in the main paper.

Added a sentence detailing that the hours selected are based on the time of a developed boundary layer, maximizing the likelihood that the tower measurement is representative of the methane in the entire boundary layer column, and added reference to Figure S8 in the supp.

Line 60: Was any filtering or data selection carried out on the observations before or after averaging? If so, please explain this process here.

There wasn't. There's a whole slew of calibration performed when collecting the observations which are discussed in manuscripts specific to the design and implementation of each basin's tower network (Monteiro et al., 2022 and Miles et al., 2018). If we want to be really specific, data during instrument or valve malfunctions and during on-site testing were excluded from the calculation of hourly averages. I've added that to the paper .

Line 67: Please add the exact latitude/longitude bounds of the two study areas, as this is not clear from Figure 1. This information would allow for easier comparison with future studies.

Added bounds to the sentence.

Line 84: Please explain what 'analysis nudging' means here.

Added an explanation and citations prior to the sentence providing simple definitions of analysis and observational nudging.

Line 114: As only 55% of prior emissions from the Marcellus study area are from O&G sources, what are the other main anthropogenic sources in this region? Are there likely to be any non-anthropogenic methane sources in the region (such as wetland or fresh water sources) that have not been accounted for in the EPA gridded inventory of anthropogenic sources? If these sources are non-negligible, please note any assumptions made about these sources and whether this may have any impact on results from this study area.

Grumble grumble PADEP. So the 55% O&G sources is not a confidence-inspiring value to be able to neglect other sources. Of course, the reason it's only 55% is because the PADEP's inventory so poorly represents O&G that it artificially makes it seem like a non-dominant source in the region. It's more likely that O&G represents ~80% of anthropogenic emissions in the domain. In truth, I maybe should have written this manuscript using the Production-based prior as the main prior for the Marcellus, but I think the argument is stronger that the PADEP inventory is wrong when starting from it as being the truth and watching everything break down in the results.

With that said, that's certainly not a compelling scientific argument to neglect discussing the non-O&G sources, is it? Alright! Here's what I've done:

-Added sentences detailing the non-O&G anthro sources in the domain (it's mostly two landfills not co-located with the O&G).
-Added a figure in the supplement mapping out the non-O&G anthro sources. I did this for the Delaware as well to be thorough, although at ~1% other sources, that may be more for comedic effect.
-Added a couple sentences on the decision to neglect wetlands in our prior. The mean WetCHARTs ensemble has it at about 3 Mg/hr (O&G is probably somewhere around 20 Mg/hr in truth, and 7 Mg/hr in the PADEP inventory). There's admittedly no easy way to test for the influence this would have on our results given the coarse resolution (our study area

fits only 6 grids of the WetCHARTs map) and huge uncertainty in wetland maps, but there's also no reason to suspect this 3 Mg/hr perfectly distributes itself in the production basin such that it biases the total in any meaningful way.
-Added a "Solving all sources" to the sensitivity analysis, where we allow the inversion to optimize non-O&G sources as well, and then extract the change to O&G by taking the total change in a grid and multiplying it by the fraction of O&G emissions that was represented in that grid from the prior. This scenario performed similarly to the default, with an emission total 8% smaller using the PADEP prior, and no change using the larger Production-based prior. This information has been added where appropriate in the manuscript, and the sensitivity analysis table has been expanded to include its results.

Line 124: Was an alternative prior for the Delaware basin also created using the production-based method that was used for the Marcellus basin prior? If so, how did this prior compare to the satellite-based inversion prior for the Delaware basin? You could comment on the choice of using the different alternative priors here, even if this was just due to differing data availability for the two regions.

The choice was mostly an availability thing. Previous work from Zhang 2020 had led to the development of multiple, smartly-constructed priors in the Permian that could be utilized for this study, whereas the northeastern Marcellus has been mostly neglected up to this point (which is rather fascinating considering it's probably the lowest leak basin in the US, maybe the world. We should be interested in that!). The question you ask may be better phrased as "why did you use a stupid production-based inventory for the Marcellus". Production-based emission rates are a bit of a flawed way to design a smart inventory (and frankly, a flawed way to report emissions), since production per well, age of well, and the methane content can create large differences in emission-normalized loss rates. For example, a conventional well in southwest PA averages a loss rate of 15% whereas an unconventional is ~0.5%. We are fortunate that in the northeastern Marcellus, the wells are all very similar due to being drilled all at similar times during the Marcellus shale boom: new, ridiculous production, all gas. It makes a simple assumption like a leak rate work as a decent substitute to a well-constructed bottom-up inventory. I've added a statement to the paper justifying the use of a production-based inventory in the Marcellus.

As an aside, a production-based inventory would look pretty awful in the Permian. There's a large mix of new and old wells with production values varying by multiple magnitudes, and lots of additional processing steps due to a mix of both wet and dry gas. It could be a cool experiment to see what one would look like, but there'd be little justification to use it since there are maps available there constructed on actual, site-level measurements.

In the northeastern Marcellus, all of the wells are of similar characteristics; new wells, ridiculous production, dry gas. In the Permian you have a variety of gas wells and oil wells of varying ages and productions of each such that applying a production-based emission rate wouldn't really make sense.

Line 172: It is stated in the supplemental that the 'hybrid filter' method of background site selection was the most successful, but that this would filter out many of the observation days, when there are complex background conditions. You could comment on how not filtering these days out, when the transport model may be underperforming due to the complex conditions, could potentially impact modelled emissions for that month.

Added sentences expressing that the elimination of data should not necessarily be viewed as a negative and that including days with less confident backgrounds can adversely affect the accuracy of the solution, but that in this study both methods produced similar posteriors.

Line 184: What is the resolution of the posterior emissions (the x terms), in units of latitude/longitude degrees, and was this the same as the prior resolution of 3x3km?

The posterior is the same resolution as the prior, in this case 3x3 km. Added clarity to that in the manuscript when describing x.

I guess I could state the unit in lat/lon degrees, but will hold off unless you really want that. I think I understand why you're requesting it; typically inventory grids are constructed in intervals of constant lat/lon. In the case of WRF, the grids are constructed in intervals of 3x3 km (every lat and lon is unique), so describing the resolution as 3x3 km probably makes the most sense.

Line 194: Were the inversions run with errors included for the other non-O&G sectors and did this have any impact on the posterior O&G flux estimates? Please state why you are able to make the assumption that the prior non-O&G fluxes have no uncertainty, particularly in the Marcellus region where 45% of prior fluxes are non-O&G.

This has hopefully been addressed better now in the changes made to the inventory description above, as well as the new addition of the sensitivity analysis allowing for the change of non-O&G sources. In hindsight, it would have made sense to just adjust all sources from the start, but other sources were mostly an afterthought until I downloaded the PADEP inventory and it lowballed all the gas by a factor of 4.

Line 313: The result of consistently lower posterior fluxes than the prior from the region with mostly newer activity in the Delaware basin is an interesting result, and could be further emphasized in the conclusions, if similar works have also found that emissions from regions of newer activity are being overestimated.

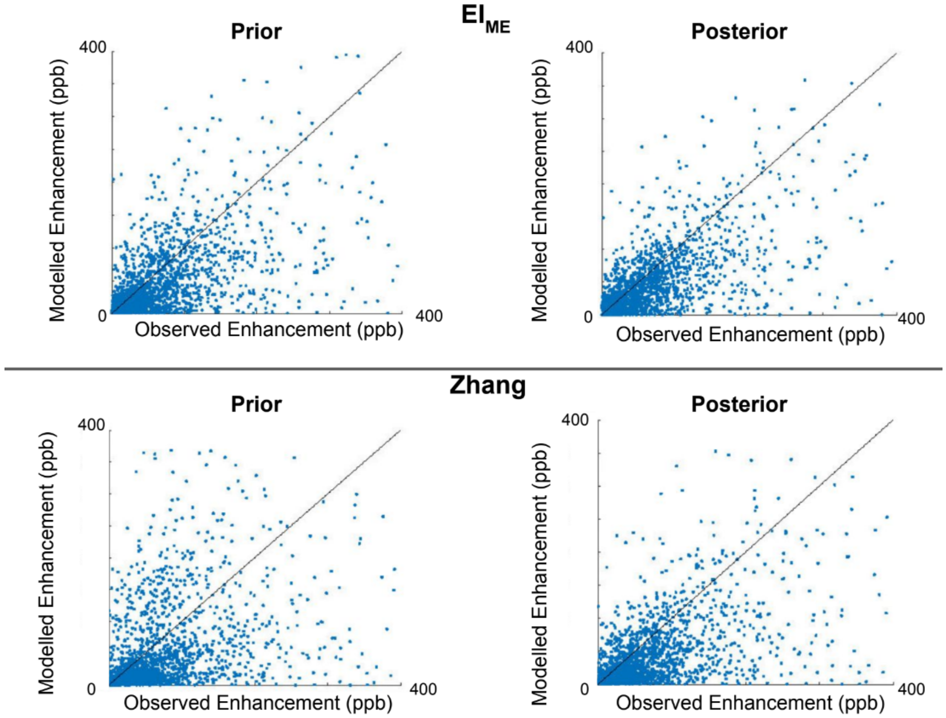
I don't think it can quite be phrased as if these emissions are being "overestimated" as the prior in this case was not really designed as any sort of official product, but rather just an extrapolation of some site level measurements to provide a basic first guess for an inversion system to then work off of. In fact, the truth is probably the opposite; the EPA's bottom-up inventories are almost entirely component-based rather than production-based, which tends to result in a larger underestimation of emissions in high-production areas (despite these higher production wells having lower production-normalized loss rates).

With that said, I can (and have now) added in the conclusions that production-normalized emission rates tend to be lower in areas with high production. Definitely not a new find, but consistent with studies such as Omara 2022.

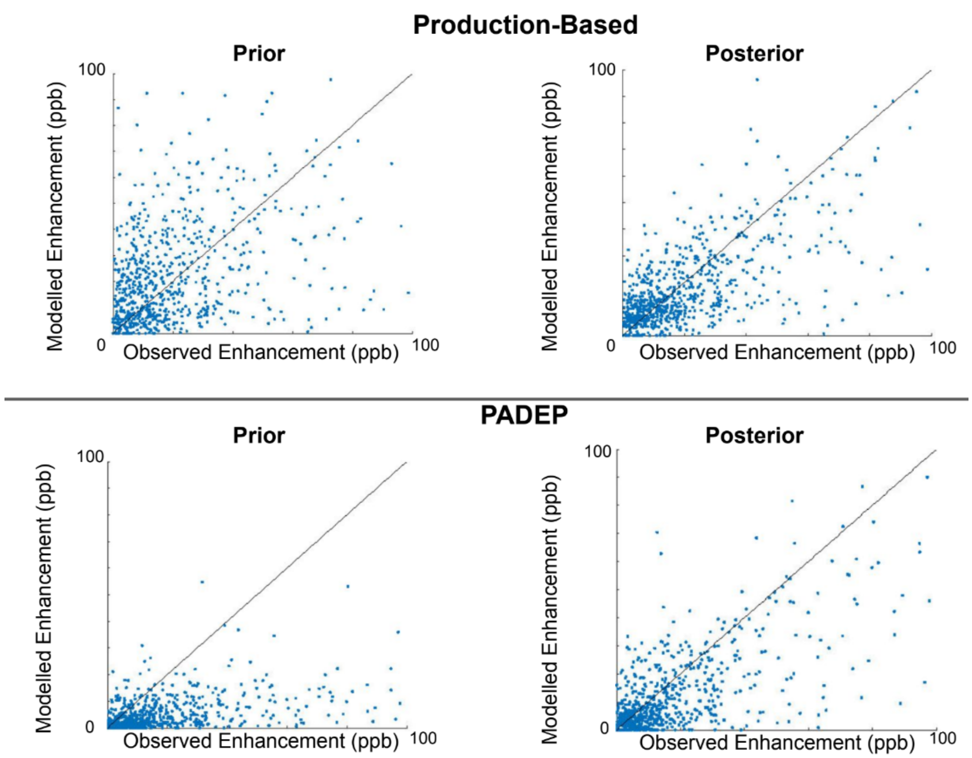
Line 338: This paragraph and Table S2 suggest that model is not fitting well to data for Delaware, or at least that the model's fit to the observations does not move the total emissions estimates from the prior, despite some changes in the spatial distribution of emissions. An additional figure, showing the difference between the modelled observations made using the prior and posterior emissions estimates could simplify the discussion of 'model-obs bias' presented in this paragraph.

There's actually a scatterplot in the supplemental that does a nice illustrative job at showing the relatively small impact of the prior vs posterior on the modeled vs observed enhancements. I just completely forgot to reference it in the main text. Oops. Reviewer 1 did a good job nagging me on the supplemental figures, and now they're all ordered properly and referenced when appropriate. I've also added a sentence early on in the inversion discussion directing the reader to the supplemental figures and tables relevant to the performance of the posteriors.

Below is the prior vs posterior obs vs model comparison for the Permian.



Here’s the Marcellus for comparison (also in the supp)



You can actually spot a bit of tightening to the 1:1 line in the Delaware plots, but the biggest thing the Delaware has going against it is that the prior's magnitude was nearly identical to what the posterior thinks it should be, so gains are harder to come by compared to the Marcellus inventories. The bias still remains a bit of an oddity though.

Interestingly, Varon et al. 2023 (also in ACPD right now) used the towers as a validation for his satellite inversion, and although his posterior does indeed improve the situation over the prior, he's still left with a 30 ppb low model bias as well. I didn't bring that up because this manuscript was supposed to be published before his was out, but given the delays, I'll go ahead and add that in the discussion.

Additional Changes

-Added a sentence to the conclusion emphasizing that intermittent sources inherent to O&G activity do not impact the tower network's ability to derive monthly emissions, as demonstrated in the manuscript.

Technical corrections

Line 56: Please define AGL here.

Defined

Line 59: The use of different notation for times is a little confusing. Does 20 UT mean 8pm Universal Time? If so, this could be corrected to read: 20:00 Universal Coordinated Time (UTC), 13:00 Central Standard Time (CST).

Adjusted accordingly

Line 65: See comment above about use of different time notation.

Adjusted as well

Figure S7: Please correct the figure description. I think this should read 'used in this study for the Marcellus basin'.

Reviewers who take the time to closely review the captions of supplemental figures are heroes. Thank you for catching this.

Some results sections are long and may benefit from being split into smaller subheadings to improve readability, (for example Sections 3.2 and 3.4) but this is at the author's discretion.

Hmmmm. I do agree the “[basin] Inversion Result” sections are somewhat long. Going through them though, each paragraph mostly discusses a different focal point than the previous one. It’d be difficult to break it up into subsections without giving every paragraph its own subsection, and then it’d probably look even more cluttered than just leaving it as is. I think I’ll keep it as is for now.