Review of Mathew et al. 2025

https://egusphere.copernicus.org/preprints/2025/egusphere-2025-1977/

General comments

Summary of the study

This study is about estimating methane emissions from India by optimizing bottom-up emission estimates using TROPOMI data in an inverse modeling framework.

In section 1 (introduction), the significance of the research question, the state of research, knowledge gaps and goals of the study are explained.

Sections 2 and 3 introduce data and methods: the datasets analyzed in this study, the atmospheric transport model, and the prior emission datasets, the sampling method and the inverse modelling framework for emission estimation.

Section 4 (results and discussion) is organized as follows. Section 4.1 describes the spatiotemporal distribution of emissions in the bottom-up datasets introduced in section 2. Section 4.2 describes the spatiotemporal distribution of XCH4 modeled using the bottom-up datasets and a boundary condition dataset. Section 4.3 and 4.4 add comparisons of the modeled mixing ratios to observed XCH4 from TROPOMI and surface CH4 from one ground-based site (Thumba), respectively. Section 4.5 shows the inverse modeling results.

The study concludes with section 5, which states the conclusions drawn from the analyses.

Overall approach

Overall, the study is a contribution towards an important goal, namely improving the understanding of atmospheric methane emissions from India. The methodological approach (inverse modeling of atmospheric methane transport to optimize a prior flux estimate) is well established in the literature. That and the comparisons to similar results from the literature provide confidence that the results are sound.

Selection of shown results

The study puts a big emphasis on describing the spatiotemporal distribution of methane emissions from published bottom-up datasets (section 4.1, 4 pages), as well as their impact on XCH4 (section 4.2, 2 pages). In my opinion, the study (abstract/introduction/conclusion) should explicitly state that such a review is one goal of the paper. Because for the goal of the study as stated in the title, "to improve methane emission assessments" using an inversion, the level of detail of these sections is too much, in my opinion. At the same time, the statements in section 4.2 and 4.3 about the separate contributions of the different sources (anthropogenic, biomass burning, wetlands) to XCH4 would actually benefit from a bit more detail, namely showing separate plots of these contributions and of the background (details in the specific comments).

By contrast, the section on emission estimation results - which should be the major part of a study that aims "to improve methane emission assessments" - is very short (section 4.5 - 1.5 pages): The only result given is the annual national total emissions and their uncertainty, and a comparison to the corresponding bottom-up estimate. Many questions one would expect to be addressed - and that could be addressed - using the results of the already performed inversion are not addressed. I recommend that the authors show more details of their results in section 4.5, i.e. the spatiotemporal and sectoral distribution corresponding to their state vector resolution, and whether/how much the fit to the observations (TROPOMI and Thumba) improves after the optimization. If there is a good reason why the authors wish to not show such additional details, it needs to be stated in the manuscript, and the expectations of the reader should be managed, for example by adjusting the title (e.g. by switching to a wording with "_towards_ improving methane emission estimates").

Language

A general language improvement is required:

- The word "the" is overused. Some, but not all occurrences, are given in the specific comments below.

- In many cases, the wording is unclear to me or appears to be inaccurate. In many cases where I found the wording confusing, it may be due to the usage of comparatives (“more”, “growing”, “enhanced”, “increased”) when positives (“large”, “high”) seem to be the appropriate choice to me. Unclear wording is what a lot of the specific comments below are about, and in many cases, I added what I think the statement should rather be.

Specific comments

Abstract

- Line 5: Please adopt the correct definition of "WRF" throughout the document ("Weather Research and Forecasting" model)

- Line 7: "non-negligible uncertainties" - should rather be something like "non-negligible differences", no?

- Line 8-9: "We find that the WRF-GHG simulations overestimate the XCH4 and underestimate the near-surface CH4 distributions".

- This sentence overstates the representativeness of the single ground-based measurement site used in the analysis. It is even a point later in the conclusions that CH4 emission estimation in India would probably benefit from expanding the observation network. The authors should harmonize abstract and conclusions on this point.

- Also, since this observation is so prominently placed in the abstract, I expected an explanation or at least a discussion of this observation. For example, it could hint at errors in the vertical transport, but also simply reflect the different representativeness of the column- and ground-based observations used here. Neither was investigated. Instead, the conclusion section only states (line 423-424): "Our high-resolution model is capable of capturing surface CH4 variability, especially for the well-mixed conditions, as confirmed by the ground-based

CH4 observations". Thus, the significance of these differences stated here in the abstract remains unclear, and abstract and conclusions need to give the same message on that point.

- Line 9: I'm not familiar with the term "first-order inversion". So I think it should be explained in the text, replaced with a clearer term or removed (the method used is an “inversion” - I see no need for an extra qualifier like “first-order”). After reading the manuscript, I think "first-order" refers to the fact that the state vector is rather coarse.

- Line 10-11: "showing that the current global emission inventories overestimate CH4 emissions considerably" - please state which inventories and by how much they overestimate the emissions.

Introduction

- Line 19: Please cite a more recent reference about recent atmospheric greenhouse gas burdens

- Line 21: If possible, please cite a more recent reference about the lifetime of atmospheric methane

- Line 28-29: This statement on OH trends is not accurate. First of all, contrary to what is stated here, the cited reference Stevenson et al. 2020 concludes that OH _increased_ recently (from 1980 to 2014). But also, as far as I'm aware, there is no definite consensus on the OH trend. E.g. Thompson et al. 2024 see no trend and cite studies with opposing results (https://acp.copernicus.org/articles/24/1415/2024/). There is also more context in Saunois et al. 2025 (https://doi.org/10.5194/essd-17-1873-2025) and Zhou et al. 2024 (https://doi.org/10.1088/1748-9326/ad4b47).

- Line 50-51: Please clarify in the text what you mean by "inadequate modeling systems"

- Line 52: "imminent" is not the appropriate term here

- Line 70-71: "We demonstrate the advantage of ... operating at a high resolution comparable to TROPOMI, which may minimize forward model-related uncertainties in the carbon assimilation system over India". This sentence should be rephrased: it suggests that the authors experimented with varying transport model resolutions, which is not shown in the manuscript. The model also doesn’t have that much higher resolution than others cited here (e.g. Ganesan et al. 2017 ran at 12-16 km), and the study does not address whether the forward or inverse model results shown here are superior to those of previous studies.

- " Section 4 discusses the results of the study from the data analysis conducted" – I’d remove “from the data analysis conducted”

- Line 71: WRF definition (see comment above)

Section 2

- Line 106: Please add the citation for ERA5 here (or cross-reference to Table 1)

- Line 110: Provide a citation, link and version of the CAMS product used here (or cross-reference to Table 1)

- What's the model top of the WRF simulation (according to Thilakan et al. it’s 50 hPa)? Please add this to Table 1. How do you extrapolate the model profile to the top of the

atmosphere to compare with TROPOMI? I.e., does your comparison method account for the CH4 decrease in the stratosphere?

- Line 112: "far-field (background)": I find the term "far-field" unclear, as it can mean different things depending on context. I suggest to remove the term "far-field" and only refer to "background" or "initial and boundary conditions". Here and in Section 4.2. Similarly, the term "regional sources", used elsewhere in the manuscript, should be properly defined (along the lines of "sources within the simulated domain")

- Line 123: Kretschmer time profiles: mention what does this entail - e.g. month + week day + time of day?

- Section 2.3 - Thumba validation station: Add something about the area it represents or at least acknowledge that it's not representative of the whole of India. "To assess the model's performance at the surface level" is overstating the representativeness. I found something on that in Uma et al. 2024, you could briefly summarize those findings here.

Section 3

Section 3 only has one subsection, section 3.1. In my opinion, the content of section 3 belongs to the methods, i.e. I would integrate it into section 2.

- Line 162ff: "The quantities to be optimized, represented by the state vector x with n elements x1 ,x2 ,x3 ,...,xn correspond to the total monthly emission information averaged over each political state in India." The term "total monthly emission information" appears to be inaccurate, because (1) wetland emissions were not optimized (line 182ff) and (2) the EDGAR and GFAS fluxes were optimized separately. Or not? Please adjust the text so it becomes clear what exactly was optimized.

- Line 170: F(x) is not only WRF-GHG but also what is sometimes called "flux model" (in your terminology: mapping of x to phi) and here, the mapping of WRF-GHG's output to total columns (Eq. (1)). Please make this clear in the text. Suggestion: Write out the relationship between phi (emission field) and x (state vector element) as an equation.

- Line 177: "the perturbed emissions"- remove "the", I thought I had missed the introduction of the perturbation. Also, specify how you perturbed the fluxes. From the equations I infer: You constructed response functions for each state vector element by adding a small flux to the prior of the state vector element. Or was it a multiplication by a 1 + a small number?

- Eq. (4) - Define "TR"

- Eq. (5) Error in the last term ("x_A" should be "x")

- Line 182ff: Earlier, the authors write that the prior represented anthropogenic, wetland and biomass burning emissions. Here, it sounds like only anthropogenic and biomass (add "burning") emissions were optimized, not wetlands. I remarked on line 162 that this it needs to be clarified what exactly was optimized separately. In addition, the authors need to state why they can omit the optimization of wetland emissions, i.e. neglect their uncertainty that was cited in the introduction.

- Line 185: I would suggest to choose one term for fluxes inside the model domain and define it properly. Both "regional" and "local" is used in different places in the manuscript, and I wondered if different things are meant.

- Eq. (9): summation symbols missing

- Line 213: Here is stated that Edgar v8 is used, earlier it was v7 - which is it? Correct all occurrences.

- Line 214: Remove "the" from "the emission activities"

- Line 220: Remove "the" from "the wetland" and "the agriculture"

- Line 220-221: Please clarify the sentence "The seasonal sources of CH4 include agriculture (see Fig.2 (b)) and biomass burning". I presume you mean something like "The sources with significant seasonality include agriculture and biomass burning"?

- Line 222ff: "The positive anthropogenic trend in CH4 emissions in India can be expected owing to the large cattle population and agricultural activities (especially rice production and waste management) as indicated by 2010-2015 GOSAT observations (Maasakkers et al., 2019)". This sentence is confusing. Do you mean to say something like “A large anthropogenic contribution to the total CH4 emissions in India can be expected …” ? The word “trend” usually refers to multiannual trends, but is used here to describe a figure that shows one year of results. Also, a “positive trend” is not explained by “large” agricultural activities as stated here, only if they were “growing”. Please clarify.

- Line 229: Fix this reference for Edgar 4.2; in the text it's cited as "Commission et al.", and in Line 514 it shows up as "Commission, E. E. et al."

Line 232: State which months are the "summer monsoon season". EDIT: Later in the manuscript, I also got confused which months are defined as winter. Please add season labels to Figures 3 and 5.

Figure 1: It's hard to see the delineation of the regions (IGP, ...) because at the boundaries, only one of the lines can be seen. Please modify the figure so that the colored borders are visible for all boundaries. Maybe dashed lines work.

Figure 2:

- Add the region boundaries as in Fig. 1.

- Why not include a panel for the coal mining sector? The authors even discuss a hot spot from coal mining (Sect. 4.1), but it's not in any figure

Results and discussion

Section 4.1

- Line 233: "the least" -> "the minimum"

- Line 233: I find the usage of percentages from here on out in this section confusing: are they the share of monthly total fluxes? Or perhaps shares of the flux in a specific sector of the annual flux of that sector? The text seems to be about seasonality of the fluxes, but if the stated numbers are percentages of the monthly total fluxes, that seasonality is not necessarily the same as the seasonality of the fluxes. Please clarify.

- Line 247 and 248: Define the regions (here: "SI" and "NI") upon first use in the text (I know they are defined in a figure caption, but as a reader, I don't want to go searching).

- Line 253: Please state that these are cities.

- Lines 252ff: I can't find these hotspots on the maps. Can you mark them?

- Line 256: "has the issue of ... possibly generating more volatile gases": First, reformulate "has the issue of" in a neutral way. Second, "More volatile gases" compared to what? Do you mean "Large amounts of volatile gases"? Third, if correct please change from "volatile gases" to "methane".

- Line 257: "increased agricultural rice emissions" - increased compared to what? I guess the authors mean "high" emissions?

- Line 271: Define "NEI" (first usage in the text)

- Line 272: an order "of magnitude"?

- Lines 278-286: These considerations are the motivation for doing the present study. In my opinion, they belong in the introduction.

- Lines 282f: "Though inverse modeling can offer better scope in updating the CH4 budget, those approaches are limited by the insufficient coverage of mixing ratio observations and inadequate representation of the process and spatiotemporal distributions in the forward models."

- I think this sentence is missing a causal connection. And it's unclear to me what "can offer better scope" means. My suggestion would be: "Though inverse modeling can improve the CH4 budget, its ability to overcome biases in the bottom-up models used as prior estimates can be limited by insufficient coverage of mixing ratio observations".

Section 4.2

- Line 289ff: Throughout this section, it's not clear to me which sources inside the model domain are considered for this XCH4 analysis: The title suggests it's only the "anthropogenic" sources, i.e. excluding biomass burning and wetlands, according to the terminology established in the methods section. However, this sentence here seems to imply that also the other sectors (biomass burning and wetlands) are included. In Fig. 4, it's yet something else: anthropogenic + biomass burning. Also, sometimes "anthropogenic + biomass burning" seems to be meant when "anthropogenic" is written (I may be wrong about that one). I get that biomass burning is at least partly from human activity in the study region, as explained a few sentences below. The text reads like biomass burning is mostly or completely anthropogenic, but that needs to be stated. Please clarify this section, e.g. by sticking to the terminology of the different sources as established in the methods section.

- Lines 289ff again: Also, if the XCH4 from wetlands is indeed not shown in this section: it's stated in lines 327-329 that the wetlands do make a contribution to the XCH4. So this needs to be shown in a figure.

- Line 290: "biomass" -> "biomass burning"

- Line 291-293: The sentence "The IGP and surrounding regions exhibit significant XCH4 enhancements (>60 ppb) from background contribution (far-field fluxes) attributed to anthropogenic and biomass-burning fluxes (see Fig. 4)" has multiple issues:

- In the previous sentence, the authors write that they discuss XCH4 enhancements due to regional sources here. Therefore, I think they meant to say "regional sources" instead of "background contribution (far-field fluxes)" here as well, otherwise the sentence doesn't seem to make sense.

- What is the significance of the number ">60ppb"? Nowhere in Fig. 4 or 5, which show the enhancements described here, does this look like a special value - it's the upper limit of the colorbar of Fig. 4, but seems to be an arbitrary cut-off. Why not state the exact numbers like "up to ..." or "from ... to ..."?

- The sentence "Seasonally, compared to summer, the highest XCH4 enhancements occur during winter (with a maximum of ~ 251 ppb in January)" and Fig. 4 or 5:

- The maximum of 251 ppb is printed as "max" in Fig. 4 for January, and can presumably be seen as an outlier in Fig. 5. So I assume the two figures show the same "max" (it's not stated what the "max" is - is it snapshots as reported by WRF? Hourly averages? Something else?). However, in Fig. 5, there are even higher outliers in Oct-Nov, whereas the "max" in Fig. 4 among these two months is shown as 142.2 ppb. So it seems that Fig. 4 and 5 are inconsistent.

- Line 293: "Seasonally, compared to summer" sounds confusing. I guess "compared to summer" can be deleted.

- State to which region the statement refers to - I guess all of India?

- The sentence "The minimum enhancement is seen in the monsoon time, possibly due to large-scale mixing in the atmosphere" has multiple issues

- What does "large-scale mixing" mean? It's not a term I'm familiar with and it's not explained in the manuscript. I guess higher wind speeds? Or do you mean higher boundary layers? Vertical transport alone doesn't affect total column measurements much, so this can't be the full story. Perhaps higher boundary layers in combination with higher wind speeds in upper layers. Please clarify. Also, please remind the reader of the seasonality of the fluxes here.

- Specify for which region this statement is true! In Fig. 5 I can see that it's not true for South India, which is affected by monsoon. So I don't understand the given explanation.

- Line 300: The phrase "high spatial distribution" doesn’t make sense – please adjust.

- Fig. 4:

- This figure doesn't convey much useful information: all the panels look the same, and only the numbers in the panel titles are referred to in the text (and they also show up in Fig. 5). Why not show an annual mean map here, and move the monthly panels to the supporting information.

- Fig. 5:

- No outliers are plotted for the South India region. Is that correct?

Section 4.3

- Line 308: Do the authors mean "reasonable _amounts_"?

- Line 310f: Please clarify "enhanced". E.g., provide context for the term "enhanced": "enhanced values ... compared to ...". Or change to a neutral term like "highest mixing ratios".

- Line 311-313: As explained above, I think a shallower vertical distribution of emitted tracers in shallow PBLs alone should not cause large gradients in total columns from TROPOMI, as its sensitivity is rather (though not completely) uniform across the lower atmospheric levels.

- Line 313: The use of the term "increased" surface emissions is unclear to me: October-November is not the seasonal emission maximum, according to Fig. 2b.

- Line 313f: "These increased local emissions, also seen in Fig. 3, are typical for some parts of India during the October-November season". State explicitly which parts of India and which flux components are meant here. The reference to Fig. 3 is not enough, because it's too much information to process. And in section 4.1, where seasonality is discussed, I don't see a statement about an October-November maximum either. I guess this statement might refer to biomass burning emissions in the IGP region, as mentioned two sentences later in context with aerosol load (still, state what is meant here) - however, their contribution to total emissions is tiny compared to the anthropogenic emissions, according to Fig. 3. So from the data presented, it's unclear to me whether there actually is an emission peak somewhere in this season – so it remains unclear to me what causes the large XCH4 values in Fig. 6 that is referred to here.

- Line 314: "... the likelihood of increased aerosol loads, ..." -> "the likelihood of bias in the observations due to increased aerosol loads, ..."

- Line 322f: "We find that the WRF simulations generally overestimate the total XCH4 mixing ratio over the Indian region compared to TROPOMI observations, peaking in winter months (>1850 ppb) (see Fig. 7)."

  - The choice of highlighting the number ">1850 ppb" is unclear to me. Before, it was stated that the mean observed concentration went up to 1870 in that period, Fig. 7g states 1883.2 ppb, so why cite 1850 ppb for overestimation by the model.

  - Clarify: do you mean that the WRF results peak there or the difference between model and observations?

- Line 324: I stumbled upon the phrase "suggesting the increased impact". Perhaps "suggesting a large impact" is meant. However, statements on contributions to XCH4 by sector (anthropogenic, biomass burning, wetlands) could be made much better when showing plots of all three (as suggested in my next comment).

- Line 327-328: "Similarly, high XCH4 values observed along the eastern coast can be attributed to the wetland emissions, as seen in Fig. 2f." I have several comments about that statement:

  - In my opinion, the eastern coast doesn't stick out with particularly high observed XCH4 in Fig. 6. Perhaps in October and November, but not really in April and May. Clarify to which season/months the statement applies.

  - Fig. 2 shows that "agricultural soils" have, on average, larger emissions than wetlands along the eastern coast. Therefore, I expect higher contributions of this sector than from wetlands to XCH4 in that area. So the conclusion by the authors that XCH4 peaks there are explained by wetland emissions is, in my opinion, not convincing.

  - If the wetland emissions had a big influence on parts of the observational dataset, it would be unclear why they were excluded from the optimization. In lines 396-398, it is even

stated that the influence of wetland emissions was very small overall - contradicting the statement here.

- There are a few options how the authors could address this comment. They could clarify which emission dataset has the higher influence on XCH4 on the eastern coast by running them as separate tracers in the WRF model, and show wetland XCH4 contributions analogous to Fig. 4. The abovementioned lines 396-398 suggest that the authors already did run separate tracers, so adding these plots to the manuscript or supplement would not be a lot of work while adding a lot of value to this section. Or they could clarify which months they mean, and state that the "agricultural soil" dataset could fit the observed XCH4 distribution too - or explain why that is not the case.

- Line 338: I feel like it's a bit of a stretch to classify October as "winter", and October - February are not classified as one season in Fig. 3. As I suggest in another comment, please add labels for the seasons to the plots.

Figure 6 vs Fig. 7:

- It looks as if the data coverage is different for the observations than the simulation. E.g. in February over the Tibetan plateau, it looks like there are much fewer data points in Fig. 6 than in Fig. 7. Was the data selection of model and observations identical for their comparison? Is it just that data points below the lower bound of the colorbar are shown in white? In that case, choose a different color than white for the lower bound to distinguish from missing data.

Section 4.4

- Line 350: This sentence is unclear: "Monthly averaged simulations and observations are found to be highly correlated in October despite their factor of magnitude differences (Figure not shown)."

- Line 355: Here, the performance of the model is evaluated against observations in terms of total mixing ratios. The fact that the model captured "79% of observed variability" is dominated by the seasonal cycle of the boundary conditions, so is not related to the accuracy of fluxes in the domain. It could be more informative to do this comparison after subtracting modeled boundary conditions from both datasets (although I acknowledge that the boundary conditions may have biases too).

- Line 360-366: Here the authors cite three possible mechanisms for the low variability and means of the mixing ratios during summer: (a) enhanced vertical mixing, (b) influx of clean air and (c) OH oxidation. The vast majority of the joint effect of (b) and (c) together could be gleaned from the contribution of a background tracer (i.e. CAMS). Hence, the suggestion in my previous comment, i.e. to separate the influence of the background, could provide insight here as well.

Section 4.5

- Line 377f: "Incorporating biomass burning emissions from GFAS has an impact of +0.3 Tg yr-1 on both prior and posterior emission estimates over the Indian region". Please clarify: Does this mean that there were two simulations, one in which the GFAS emissions were held constant and one where they were optimized?

- Line 387: Which inventory? Also EDGAR? What were the estimated emissions?

- Line 388: "improve CH4 emissions in India" -> "improve CH4 emission estimation in India"

- Line 391-392. "Prior emission errors in the spatial distribution, which would have propagated from data sources or bottom-up methodology, cannot be minimized in our optimization procedures". I disagree. They were optimized, just with coarse resolution.

- Line 393-395: This closing sentence has nothing to do with the rest of the paragraph - please reorganize.

- Lines 396ff: In my opinion, the statements on the various emission estimates in the literature belong in the introduction.

- Line 396-398: "We excluded natural wetland emissions from the inverse optimization as we focused on major anthropogenic emission sources. Further, the natural wetland prior emissions have resulted in negligible impacts on the column mixing ratio enhancement (figure not shown), which is smaller than the uncertainty of the satellite measurement." I would remove the first sentence - it's not a justification why excluding wetlands from the optimization would be reasonable in this case. Also, as I suggested above, it would add a lot of value to show this figure. Finally, the statement contradicts the above discussion of how wetlands allegedly have a noticeable influence on observed XCH4 at the eastern coast (see my comment on that above).

- Line 399: "However, Indian wetland emissions also vary based on inventories and prior models" - unclear. Do you mean they "vary among bottom-up models"?

- Line 399: "Approximately 7.5 Tg CH4 emissions from the Indian region ..." -> "Approximately 7.5 Tg CH4 _wetland_ emissions from the Indian region ..."

Section 5 (Conclusions)

- Line 418-419: "The highest seasonal enhancements of anthropogenic XCH4 occur during winter, influenced by agricultural emissions, biomass burning, and atmospheric winter transport". As stated above, I think the authors did not sufficiently demonstrate how exactly winter transport (i.e., shallow PBL height) influenced the total columns of CH4. I think they could do so by calculating mean wind speeds in the atmospheric boundary layer, which are the winds that the emitted methane is transported with, and may find that they are higher in summer presumably because the boundary layer reaches higher. Or cite a study where this was shown.

- Line 422-423: "The total XCH4 along the eastern coast suggests the effects of wetlands on atmospheric column mixing ratios." As stated above, I am not convinced that it isn't rather the agricultural soils. As also remarked above, this statement contradicts the statement elsewhere in the manuscript that wetland emissions had a "negligible" effect on XCH4.

References

- Many links follow the pattern "https://doi.org/https://doi.org/<doi>". While they seem to work, this is probably not intended.