Reconciliation of observation- and inventory- based methane emissions for eight large global emitters

Referee #1

Petrescu et al. compiled observation- and inventory-based methane emission estimates for 8 large global emitters. These data from different sources all come with inconsistent formats and sector partitions. The authors harmonized the dataset with a consistent sector partition, so different estimates can be compared properly, which is valuable. However, I have several concerned that (1) the paper contains substantial discussions that are not directly related to the dataset, which distracts the main purpose;

We are grateful to Reviewer#1 for taking time to read and comment on this draft and we are pleased to hear that, even if currently does not recommend the paper for its publication, he finds the work valuable.

Regarding the first concern (1): The paper was built on the results from the three-year EU funded CoCO2 project. Section "4. Discussion and recommendations on reconciliation procedures" shares our experiences of performing reconciliation and how this can be improved, which comes from a project deliverable on a blueprint for a decision support system to be used in an eventual CO2MVS. We believe this adds value as it builds on dialogues with a much broader community of users, e.g. scientists, inventory agencies, policy makers etc. taking into account their opinions and needs when it comes to comparisons between the two approaches.

Having said that, we will try to shorten and make the discussion more concise. If the reviewers and the editor in charge will agree, we would still like to include the section 4 discussion as an important input on user engagement guidelines to be produced in the future under different communities (WMO and IG3IS, and EU funded projects).

(2) the statement that observation- and inventory-based methane emissions are reconciled is an exaggeration and is misleading, so, I cannot recommend the paper, in its current form, to publish in ESSD. Below explains my main comments.

Regarding Reviewer's concern on the use of the word "Reconciliation" and that it may be an "exaggeration and misleading". This likely depends on the context in which the word is used. In our case we see the word as what we are doing (reconciliation) as opposed to what was successfully done (reconciled): reconciliation is the process or action of making one dataset comparable with another to assess their consistency. Reconciliation is very much the process, not the outcome.

In this respect, we tried to obtain consistent results from both BU and TD estimates, and the compatibility between BU and TD products implies harmonization of the results, concepts and definitions. As we are comparing BU and TD approaches, I believe we can, in terms of our purpose, and including the available data we have, present it as an attempt to reconcile the two approaches.

We will add a clearer definition of what we mean by reconciliation in the revision.

Main comments

The objective of the paper is not clearly stated. As a data paper, one would think that main objective is to describe the dataset. In the case of this paper, since the data are taken from other studies, the key is to describe how diverse data are harmonized and what the harmonized data tell us.

Thank you for your observation. It is true that one should explain the harmonizing procedure and for that we show how diverse the data is (Table 2 and Figure 5) and for comparability we also try our best to harmonize it (Table 3 and Figure 6). We will rewrite this section to better explain it.

However, Section 4 contains substantial discussion that is very general and not directly related to the dataset, which is distractive.

In our opinion, this section is really important and constitutes a message from a user engagement perspective. Ideally, we want close collaboration between the scientific community and the inventory community, if their emissions are to be verified. For EU Member States to be able to use scientific products to complement their inventories, monitoring and verification procedures are highly important. Key issues in reconciliation identified during our research (CoCO2) are particularly relevant when inventory agencies want to use our scientific data and analysis to complement their inventory work.

The recommendations given in the end are random and not backed up by the findings of the paper. I would suggest that the authors clearly state their objective and organize the content around that.

We thank the reviewer for this suggestion. These recommendations are not random, are the summary of Section 4 and the findings from long comparisons exercises between reported data and scientific BU and TD estimates. We agree to the reviewer's suggestion and will better link the findings to the objectives.

Specifically, in Line 134-135, the authors claimed that the paper "aims to inform and attract attention of the use of the results for diverse climate stakeholder needs beyond research", which is a very good statement of the objective. However, it was then not discussed anywhere in the method and results. It is not never explicitly explained what prevented stakeholders from using the existing methane emission data, why this dataset compiled by the authors would be more appealing, and what efforts have been made to make the data easy to use. Moreover, I checked out the data in the repository. The datasheets still look very complicated to me, and I am not sure that non-researchers can easily find the information they needed. Anyway, more explanations are needed if the above statement is indeed the objective of the paper.

Thank you for your positive comment. Indeed the objective of this paper is linked to user engagement. Therefore, we will make sure we will incorporate adequate discussion regarding the involvement of stakeholders in this process. Regarding the format and complicated data from the repository, according to the policy of ESSD, we provide the numbers behind the figures and, as previously done for other publications, we provide the readers with the timeseries of each dataset per country. We believe it ensures replicability for anyone who can use excel/python in a basic/intermediate way.

2. "Reconciliation of observation- and inventory-based methane emissions" is used as the title and presented as the main finding of the paper. I find it an overstatement. The paper presented "total inversion methane emissions > BU anthropogenic emissions" as a discrepancy, which was reconciled by considering "posterior total flux from inversions (roughly) = BU anthropogenic flux + BU natural flux" at the national scale. However, this level of consistency/reconciliation is not surprising at all. Why would anyone want to directly compare total methane emissions from an inversion with just anthropogenic emissions from an inventory and ignore natural sources? So, I feel that the title exaggerated what was found in the paper. It is very likely that a more in-depth investigation into the data would identify significant discrepancies between the observation- and inventory- based methane emissions.

As already explained above, the title includes the term reconciliation from the perspective of the process of comparing two or more sets of results, ensuring their accuracy and consistency. For

consistency, we performed and explained the harmonization procedures, regarding partitions and priors present in each dataset (Tables 2 and 3).

We also discuss not only total TD versus anthropogenic BU, but we include the natural emissions as well. There is a section dedicated to the harmonization of natural emissions from TD partitions (Table 4). In this respect, the title talks about BU and TD reconciliation, and takes into account all available CH4 sources. This comparison, which we believe is complete, identifies already inconsistencies which are summarized in the section 3.4.2. If the scientific community wants to assist Members States in improving reporting (e.g. improved EFs, detect and quantify accurately the hot-spot emissions, spatial distribution of emissions), and supporting the national reporting agencies in using atmospheric data to complement and complete their NGHGIs, we should be able to explain differences and take into account all existing sources. There will be always discrepancies between datasets, as they do not include the same partitions, however if we are able to emphasize and quantify them, this will clarify the aim of this work. We agree that this harmonization section is not very well structured, so we will address again these issues and give a clearer and more organized explanation.

A more important questions is the reliability of the comparison made here between the inversions and inventories. Inversions are regarded as independent top-down verification of bottom-up inventories. But they are not. All of these inversions rely on prior information and therefore not independent of bottom-up emission inventories.

The reviewer is right that inversions do not provide GHG emission estimates that are independent from the prior that is used. Because different inventories rely on different methodologies and Tiers (e.g., IPCC defaults (Tier 1) versus country-specific data (Tiers 2 and 3)), priors used in inversions are often different from values reported in NGHGIs, which already provides some independence. Furthermore, and more importantly, the atmospheric measurements do provide independent information about emissions. Therefore, the evaluation against atmospheric measurements and adjustment of emissions needed to bring the atmospheric transport model in agreement with them is based on independent information. The inversion gives an indication of how far the atmospheric data suggests to move away from the prior.

Having said that, we believe it is useful to compare the two approaches, as long as we highlight these dependences and explain the differences.

Comparing bottom-up inventories and inversions without characterizing this dependence makes it difficult to judge whether the two are actually consistent. For example, we do not know whether the agreement of the EU emission trend from various inversions was due to strong observation evidence, or due to similar prior information used by the inversions and a weak observation constraint. We also do not know if the disagreement in the USA emission trend was due to a weak observation constraint and different prior information. If this is the case, it makes little sense to talk about the consistency between the inversions and bottom-up inventories in terms of the USA emission trend.

As previously discussed in Petrescu et al., 2020 (Figure 4) regarding trends for the EU27 from BU estimates, we agree that EU trends have similarities due to prior information used by inverse systems, which are coming from the BU results.

Both BU inventories and the NGHGI use similar activity data and, to varying extents, the default EFs reported in the IPCC (2006) guidelines, meaning that the estimates are predestinated to agree rather well. Thus, the spread in all BU estimates may not be indicative of the uncertainty.

The few inconsistencies between CH4 BU estimates and NGHGI are mainly caused by different methodologies in calculating emissions as highlighted in Petrescu et al. (2020, 2021).

For the USA, we explained why the trends for the CTE and MIROC are different from those from CAMS and TM5-4DVAR (Tropomi based). This is explained in the discussion after Figure 4 (as well as in the Supplementary Information) and this is due to the fact that the priors for oil & gas emissions for CTE and MIROC are based on the GAINS model (which shows a similar trend in Figure 3) while the other models use EDGARv6.

For a good overview of differences between priors, we have in the Zenodo repository an excel file which includes a spreadsheet named "Priors" summarizing each model and the priors they use for different partitions. https://doi.org/10.5281/zenodo.10276087. We believe that in most countries, these are the reasons for differences.

Minor comments:

We thank the reviewer for these comments. Some of the comments were also highlighted by the Reviewer in the major comments section above, and we believe we answered accordingly.

The minor issues we will address accordingly in the revised version of the paper.

Line 128-129: The statement indicates that achieving the climate goal will automatically lead to gains in areas of energy, food, etc., which can be misleading. In fact, controlling methane emissions may pose significant challenges to energy and food security. I suggest rephasing the statement to be more balanced.

Line 190: Spell out LULUCF at the first appearance.

Line 191: Missing information. "...which according to the ? are defined as... "

Line 199: Period sign before "Furthermore".

Line 241: What is IPPU? Spell it out and explain if necessary.

Line 291: e.g. -> i.e.

Line 301-302: It may be useful to report the rate of reduction in USA emissions, as a comparison to the EU value reported above.

Line 339: Perhaps be more specific that Russian CH4 emissions remained rather low "relative to its pre-2000 levels"? Compared to other countries, Russian emissions are not low at all.

Line 383: What is AD? Activity data? Spell out and explain if necessary.

Line 464-473: The paragraph appears to be out of context. All the remainder of the section discusses the BU and TD comparison (in terms of both average emissions and trends), while this paragraph talks generally about sectors driving CH4 growth.

Line 484-490: This result shows that the emission trends derived from the inversions are strongly dependent on the prior choice, indicating that the atmospheric observations used in these inversions are inadquate to constrain the emission trend. The relatively good agreement of emission trends in other countries (e.g., EU) also does not provide strong evidence, because the agreement can be driven by similar prior information.

Line 501-502: Again, this may be due to different choices of prior emissions.

Figure 6 and Line 600: Biomass burning is considered anthropogenic in Figure 6 but natural in Line 600. I understand both anthropogenic and natural processes contribute to biomass burning. However, the current description is unclear and confusing. A clear description and terminology should be given to distinguish its anthropogenic and natural components.

Table 3 and 4: The orders of inversions are listed differently in the two tables, making it difficult to compare.

Table 4: For an inversion, there is a difference between "missing" and "unreported" sources. If these natural sources are included in the model simulation but are not reported as results (for example because they are not optimized by the algorithm), it makes sense to use BU estimate in place in order to compare "apple to apple". However, if these sources are not included in the prior simulation, the posterior total flux inferred from observations may still implicitly include their contributions because the observation sees the total flux, although the fluxes from these sources can be mis-attributed to other sources. If this is the case, adding BU estimates to inversion estimates will actually lead to inconsistent comparison. Therefore, it is important to distinguish between "missing" or "unreported" sources, or discuss the complication.

Line 701-707: The discussion on city-level and even event- or facility-level inversion is irrelevent to this study. The entire paper is on national emissions. It is still unclear how information is integrated on these very different scales.

Line 723-724: Worden et al. (2023) provides a framework to properly compare inventory and observation-based inversions.

Worden, J. R., Pandey, S., Zhang, Y., Cusworth, D. H., Qu, Z., Bloom, A. A., et al. (2023). Verifying methane inventories and trends with atmospheric methane data. AGU Advances, 4, e2023AV000871.

Line 724: "Some" attempts

Line 770: I think this is a very good recommendation. However, it is inadequately discussed and justified in the paper. A reader may want to learn the justification of this and other recommendations.

We will add to each of these recommendations links to the discussion and additional explanations.

Line 738-740, Line 776: What do you mean by measurement of fluxes? Please be explicit whether you talking about regional fluxes or fluxes of specific sources? In the case of latter, it is actually emission factors rather than fluxes that are directly useful for a better prior estimate. Moreover, activity data, in many cases, are also bottle-necks for better priors and better uncertainty estimates, in addition to emission factors.

Line 922-930: Redundant reference information.

Line 832: Inversions are not entirely independent data, as they rely on the prior information.