Answers to reviewers: ESSD-2023-210

CoCO2-MOSAIC 1.0: a global mosaic of regional, gridded, fossil and biofuel CO2 emission inventories

Ruben Urraca, Greet Janssens-Maenhout, Nicolás Álamos, Lucas Berna-Peña, Monica Crippa, Sabine Darras, Stijn Dellaert, Hugo Denier van der Gon, Mark Dowell, Nadine Gobron, Claire Granier, Giacomo Grassi, Marc Guevara, Diego Guizzardi, Kevin Gurney, Nicolás Huneeus, Sekou Keita, Jeroen Kuenen, Ana Lopez-Noreña, Enrique Puliafito, Geoffrey Roest, Simone Rossi, Antonin Soulie, Antoon Visschedijk

REVIEWER #1

Useful and important work that certainaly needs to be documented.

Development of datasets where best and most recent regional information can be used and integrated into global products is obviously needed. While initially there might be some trade offs between overall global consistency and level of detail included allowing for, for example, improved spatial features, there is a long term advantage that will lead to identifying and resolving issues or quickly improving quality for the global product. At the same time, such work can feedback to local/regional developers to address some of the issues identified so that further versions will be seamlessly integrated in global work to keep it up to date. Some of the aspects how this work can serve and stimulate the process of continues improvement of both regional and global products could be reflected in the paper and conclusions.

In general, this is a very 'dry' technical text which could benefit from some simplification to highlight key elements of work and challenges when compiling the inventory while leaving some of the detailed descriptions to SI. I have been struggling a little to find key information about the data flow or process in the main text while liked very much the very well written concise information about the inventories in the SI; probably the first time where I enjoyed reading, or going through, SI more than the main paper ;-).

More specific comments:

#1.1) Line 36-37: I am a bit confused reading the sentence starting with 'Most regional inventories...'. Not sure what are the authors saying; isn't it that typically the local/regional inventories will have more information that global products and so will improve over those? Here it reads as they are compared against the global LPS database that is a benchmark?

Answer: We fully agree with the reviewer's comment. Regional inventories provide more detailed information on super-emitters than global ones: regional inventories provide large emitters as point sources (instead of gridded values) including regional information on the location of power plants and large industries, while global inventories generally don't do that. We used the CoCO2 power plant database to evaluate this potential improvement. The results corroborate this hypothesis: all super-emitting pixels from regional inventories contain a power station whereas several super-emitters from global inventories are likely incorrectly geo-located.

We have re-written this part of the abstract as follows:

"All super-emitting pixels from regional inventories contain a power station (CoCO2 database) while several super-emitters from global inventories are incorrectly geo-located, which is likely because regional inventories provide large energy emitters as point sources including regional information on power plants location."

#1.2) Line 67-74: Could authors provide a more explicit information as to what and why information may be lost due to these requirements? The following sentences provide examples but why is it that a more detailed regional information (if available) cannot be included in globally consistent inventories?

Answer: The paragraph is particularly on globally consistent inventories used for operational CO_2 inversions. For this application, global inventories need to be spatially consistent to avoid spatial heterogeneities in the inversion model. As described in the introduction (lines 69-75), there are two main requirements that may lead to the loss of information.

- Operational CO_2 inversion systems need near-real-time emissions so they cannot wait until detailed emission information (e.g., regional energy statistics) is published.
- Global CO_2 inversion models need spatially consistent bottom-up inventories not only to provide accurate estimates but also to facilitate the resolution of the inversion problem.

The following sentences provide examples but why is it that a more detailed regional information (if available) cannot be included in globally consistent inventories?

This refers mostly to gridded emissions such as those from roads, settlements, or aviation trajectories. If a more detailed dataset is available in a specific region, but the dataset is not consistent with the global dataset (both in terms of emission magnitude and the geo-location of the emissions), this would create a spatial discontinuity in the border between the two inventories that would affect the inversion model.

We have clarified this as follows:

"Second, gridded inventories should provide spatially and methodologically consistent emissions for global inversion models. This may lead to the exclusion of more detailed information available in some regions because spatial inconsistencies in the border between two regions (e.g., spatial discontinuities in road or aviation emissions) would have a negative impact on the inversion model."

#1.3) Table 2: In the third row thee is '-' for the column of CO2. What does it mean? i'd expect reference to ff, bf...

Answer: Corrected. The temporal profiles of CAMS-GLOB-TEMPO 3.1 refer to CO_2ff.

#1.4) Line 150: Could authors add a word of rationally why biofuels are included and agr and wildfires are not?

Answer: The reasons of excluding the emissions mentioned by the reviewer were:

- Wildfires: CoCO2-MOSAIC only covers anthropogenic CO_2 emissions.
- Agriculture: CoCO2-MOSAIC covers CO_2 emissions from agricultural soils ("other" macro-sector) or agricultural transport ("transport" macro-sectors). The main anthropogenic component excluded is agricultural field burning because the burning of "renewable biomass" is generally considered carbon neutral. Despite it can influence the climate for years before it is re-sequestered, CO_2 emitted in burning biomass is not considered a contributor to long-term climate change if the equivalent biomass is regrown, as in crop farming.

#1.5) **Table 3: Category 4C refers to 'waste incineration' and so i assume since this is under energy sector refers to incineration with energy recovery. Do you cover also emissions (and where) from open burning of municipal or industrial waste?**

Answer: CoCO2-MOSAIC follows the IPCC classification (see Table 3), so the waste sub-sector (4C) includes both waste incineration (4C1) and open burning of waste (4C2). The choice of including all these emissions in the "energy" sector was due to the high uncertainty of separating incineration with and without energy recovery. Therefore, for simplicity, all waste incineration emissions were considered in the energy sector. This choice was made in the predecessor project, CHE (Choulga et al 2021, https://doi.org/10.5194/essd-13-5311-2021), and was kept in CoCO2 also for consistency.

In any case, the weight of waste incineration emissions in the energy sector is insignificant. For CO_2ff, the weight is below 0.2% in all regions. For CO_2bf, the weight is above 0.2% only in Europe (2.8%) and East-Southeast-South Asia (3.3%). Therefore, the impact of this choice on the comparison against other inventories is negligible.

We have added the following clarification in Section 2.2.3:

"For simplicity, solid waste incineration includes both incineration with and without energy recovery due to the high uncertainty of separating these two groups. This choice was made at the CHE project (Choulga et al 2021) and was kept in CoCO2 for consistency."

#1.6) **Table 3 on page 8: Is this really a continuation of Table 3? It seems to have a different structure.**

Answer: We have added the IPCC sector column to make clearer that the Table on page 8 was a continuation of that on page 7.

#1.7) **Table 4: See question above on open waste burning, i.e, municipal or industrial waste. Is this covered or not in the 'waste incineration category'...I assume not but then do these inventories account for that?**

Answer: See answer to comment #1.5.

#1.8) **Section 2.2.3. The Table 3/4 are not easy to plow through and while they contain some detailed information, I was wondering if there is a way to develop and overview table or a map that can show which (global or regional) inventory is used for a given region/sector indicating where gap-fills are done and with which inventory. Stay at a general level to give an overview rather than a lot of details. The detailed tables can be included in the SI and the won space could be possibly used to explain what challenges are still remaining, even after gap-filling so that the national/regional em inventory developers could potentially address them. Current text and information in the tables makes it, in my view, very difficult to get an overview.**

Answer: Thanks for the suggestion. We have added the gap-filling of sectorial emissions to the flowchart of Figure 1, and moved Table 4 to Supplementary Material.

#1.9) **Line 175: I realize that maybe the difference will be small but in principle, one probably can allocate the information to respective countries by splitting emissions by area within the pixel belonging to each specific country?**

Answer: Yes, the aggregation methodology suggested by the reviewer would give more accurate results. This methodology could be implemented by CoCO2-MOSAIC users with country polygons or any other spatial polygon of their interest. However, this method cannot be implemented in a unique raster layer to be provided as part of the CoCO2-MOSAIC dataset. This is why we assigned each pixel to the country covering most (>50%) of the pixel. The goal of providing a country layer was to facilitate non-expert users the aggregation of the emissions at country level.

We have clarified this as follows:

"Note that this could introduce a small error when using the country mask to aggregate the emissions per country in those countries with a significant share of their emissions close to their borders. These errors are negligible at global scale, but users could use their own aggregation algorithms accounting for the exact area covered by each country to eliminate them."

#1.10) Table 6: Suggest reducing use of 'swd' ene, Res, tro+trn+.... can these be spelled out like in other columns? These three letter codes are also not explained anywhere in the paper; i know several can be easily guessed..but not all

Answer: We have spelled out the acronyms as suggested by the reviewers. The full description of CAMS-GLOB-ANT 5.3 sectors is available at Table S9.

#1.11) Line 347: The sentence starting with 'Small differences..." Looking at the Fig 3, they seem larger than in totals?

Answer: We have rephrased as follows: *"Some differences exist..."*

#1.12) Line 390: China is typically not classified as part of 'South East Asia' but either as part of 'East Asia' or 'North East Asia'

Answer: Thanks for the clarification. We have replaced "South East Asia" by "East, Southeast and South Asia" (abbreviated "E, SE and S Asia" in the figures), which is the description of REAS spatial coverage given by *Kurokawa et al 2020*. For simplicity, we have also used the term "REAS region" throughout the discussion section.

#1.13) Lie 460 to 477: Some of these statements and discussion is referring to great details and I wonder if this can be simplified highlighting typical (occasionally occurring) features of misplaced, lost information and refer to some specific examples which can be moved to SI.

Answer: Thanks for the suggestion. The analysis of each individual false positive (potential error on power plant geolocation) has been moved to Supplementary Material.

#1.14) Conclusions: I am missing here clear statements why this inventory shall be used and is an advancement compared to other work. Further, bring in key identified gaps/problems and suggest either solutions or how further work could resolve them

Answer: Thanks for the suggestion.

Regarding why the mosaic shall be used, we have clarified it as follows:

“CoCO2-MOSAIC 1.0 could be considered a globally accepted reference that can be recommended as a global baseline emission inventory. The mosaic provides harmonized access to regional inventories at a global scale facilitating the replication of inter-comparisons such as the one made in this study.”

Regarding the gaps/problems identified:

“CoCO2-MOSAIC has been used to benchmark global emission inventories identifying the main sources of discrepancy in each sector and region, giving valuable feedback to inventory developers to continue improving both regional and global emission datasets.”

REVIEWER #2

This study contributes to global high-resolution CO2 emission inventories by integrating data from several regional inventories. The undertaken workload is substantial, and the resulting dataset is valuable. However, there is room for improvement in the text to clearly articulate the limitations of previous studies and the advancements introduced by CoCO2-MOSAIC.

Answer: The current version of the introduction describes extensively the value of collecting regional emission information for assessing the quality of global emission inventories. The main limitation of using regional inventories is that these datasets are spread in different institutions and have different data formats, resolutions, or sector descriptions. Therefore, most studies currently focus on the comparison of global vs global inventories, or global vs one regional dataset. CoCO2-MOSAIC solves this problem by providing a harmonized version of all regional inventories, so users can easily replicate inter-comparisons like the one conducted in this paper with their inventories.

We have clarified this as follows:

"Note that the use of regional emission datasets for assessing global inventories is currently limited by their accessibility (e.g., different spatial resolution, sector description, or data format). CoCO2-MOSAIC 1.0 solves this issue by providing harmonized access to regional datasets at a global scale, helping users to replicate inter-comparisons such as the one conducted in this study."

#2.1) Additionally, some errors need attention, such as unifying the terms COCO2 and CoCO2.

Answer: Thanks for pointing this out. We have reviewed the manuscript using always "CoCO2".

Here are some detailed questions and suggestions:

#2.2) REAS (Regional Emission Inventory in Asia): I think it includes East, Southeast, and South Asia, not exclusively Southeast Asia. Please check.

Answer: Same as Comment #1.12.

We have replaced "South East Asia" by "East, Southeast and South Asia" (abbreviated "E, SE and S Asia" in the figures), which is the description of REAS spatial coverage given by *Kurokawa et al 2020*. For simplicity, we have also used the term "REAS region" in the discussion section.

#2.3) L33: "CoCO2-MOSAIC1.0 has the highest CO2ff and CO2bf emissions globally…" Would it be better to provide specific numerical values or explanations for clarity?

Answer: Thanks for the suggestion. We have added the numerical values to the abstract.

#2.4) L36-L37: "Most regional inventories…" is confusing, as pointed out by Reviewer 1. Besides, large emitters are not limited to power plants.

Answer: Same as comment #1.1.

We have re-written that part of the abstract as follows to clarify this aspect:

"All super-emitting pixels from regional inventories contain a power station (CoCO2 database) whereas several super-emitters from global inventories are incorrectly geo-located, which is likely because most regional inventories provide large energy emitters as point sources including regional information on power plant locations"

Could you provide more clear information about the point-source emission in each database you use?

It is true that large emitters are not limited to power plants, but in this study, we defined super-emitters as CO_2 energy sources above 7.9e-6 kg/m2/s. The point-source database we used to benchmark super-emitters of regional and global inventories is the CoCO2 power plant database (described in Section 3.3.1).

#2.5) L85: "Compared to global inventories, CoCO2-MOSAIC 1.0 includes all available regional information, without the limitation of providing spatially and methodologically consistent emissions." You may clarify further on the limitations related to "spatially and methodologically consistent emissions."

Answer: Same as Comment #1.2.

This refers mostly to gridded emissions such as those from roads, settlements, or aviation trajectories. If a more detailed dataset is available in a specific region, but the dataset is not consistent with the global dataset (both in terms of emission magnitude and the geo-location of the emissions), this would create a spatial discontinuity in the border between the two inventories that would affect the inversion model.

We have clarified this as follows:

"Second, gridded inventories should provide spatially and methodologically consistent emissions for global inversion models. This may lead to the exclusion of more detailed information available in some regions because spatial inconsistencies in the border between two regions (e.g., spatial discontinuities in road or aviation emissions) would have a negative impact on the inversion model."

#2.6) Figure 1: If "x" means "No", why GEAA-AEI is used for Gap-filling?

Answer: We have clarified the flowchart nomenclature by using a tick (✔) to indicate those processing steps applied to each inventory. We have described it in the Figure caption as follows:

"The tick (✔) means that the specific processing step was applied to the inventory."

#2.7) 3.3.1: You may define "super-emitters" and explain the choice of flux > (7.9e-6 kg/m2/s) as indicated in Table 9.

Answer:

Super-emitters are pixels with energy emissions above 7.9e-6 kg/m2/s. The threshold was defined in a previous project, CHE (see Choulga et al 2021, https://doi.org/10.5194/essd-13-5311-2021). The goal of that study was to calculate the uncertainty of 0.1x0.1 global anthropogenic CO_2 emissions. The authors checked manually the accuracy of most emitting energy pixels (super-emitting) to reduce the uncertainty of the energy emission estimates. The choice of the threshold (7.9e-6 kg/m2/s) was made to filter a reasonable number of super-emitting pixels at a global scale that could be checked manually by the CHE team (see Figure below). We kept the same definition in CoCO2 (successor of CHE project) for consistency with CHE results.

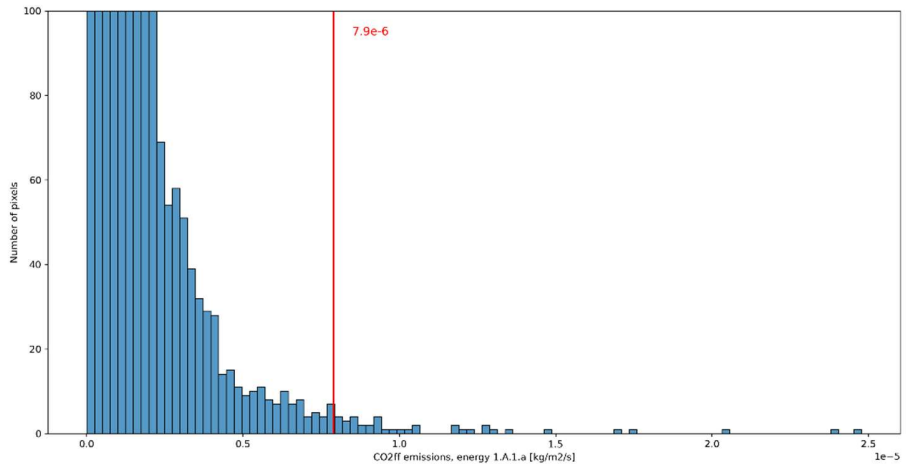

Figure 1 Histogram of Energy 1A1a emissions per 0.1x0.1 pixels

Based on the reviewer's comment, we have made the following modifications to the manuscript:

- We have included the definition of super-emitters in Section 3.3.1
- We have explained the reason for the threshold in Section 2.2.3: *"The choice of a threshold of 7.9e-6 kg/m2/s was made in Choulga et al 2021 to filter a reasonable number of super-emitting pixels whose accuracy could be manually checked to reduce the uncertainty of energy emissions "*

#2.8) L470: Despite the explanation of REAS 3.2.1's coarse resolution, the missing of three big power plants in China (Figure 6) is noteworthy. I see that your focus on power plants in another submission (Guevara, 2023), do you have an explanation for this discrepancy?

Answer: As mentioned in the document, REAS coarse resolution influences the total number of REAS emitting pixels but not the number of super-emitters because REAS provides large emitters as point sources (and are assimilated as such by CoCO2-MOSAIC).

- Regarding super-emitter geolocation, the agreement between REAS & CoCO2 power plant database is perfect (all super-emitters contain a power plant), likely due to the use of point source information. By contrast, global inventories show different geo-location errors.
- Regarding the smaller number of REAS super-emitters (and potentially missing super-emitters in China) the reason could be that REAS energy emissions are smaller than those of global inventories (Figure 3), which could be either due to smaller activity values or emission factors in REAS. This potential underestimation would lead to smaller emissions also at the power-station level, explaining why fewer pixels are over the threshold defined to classify super-emitters (7.9e-6 kg/m2/s)

#2.9) L555: As the regional inventories are available only for 2015, I worry about the continuity of emissions from 2015 to 2018. Have you conducted any analyses addressing this potential discontinuity?

Answer: The main goal of CoCO2-MOSAIC is to include all the regional information available for at least one year, so all this regional information can be used to benchmark global emission inventories. We selected the last year in which all regional inventories are available, 2015.

CoCO2-MOSAIC does not intend to evaluate the temporal trends of the emissions. Therefore, from 2016 to 2018 the goal is not on evaluating temporal trends but on providing the emissions available in some regions, so CoCO2-MOSAIC users can extend their applications in these regions if needed.

The methodology used to build the CoCO2-MOSAIC could be used in the future to extend the temporal coverage of the mosaic once the temporal coverage of all regional inventories is extended.

This aspect is addressed in the Limitations section, “temporal coverage” bullet point.