Replies to comments provided by Referee #1

Ref.: essd-2024-410
Title: **Global gridded NOx emissions using TROPOMI observations**
Author(s): **Anthony Rey-Pommier et al.**
Earth System Science Data
Type: Research article

We would like to thank the reviewers for their careful reading, that led to interesting comments. Reviews have been addressed in the revised manuscript and commented in this document. For a better readability, reviewer comments are highlighted in grey in this document, while answers are highlighted in light yellow.

This is an excellent study estimating NOx emissions in 2022 from TROPOMI data. I appreciate all the assumptions involved and acknowledge that there are several additional sensitivity studies that could be done but also understand most of them are beyond the scope of this manuscript. With that said, I have listed a several minor suggestions that could improve the paper and better clarify some of the unstated nuances of the work.

Major comments:

It's unclear exactly how OH is being incorporated to estimate the NO2 lifetime. Are you using surface OH concentration at the closest CAMS grid point? Or a model weighted vertical average based on the NO2 distribution? Or something more technical? If it's the former, I recommend authors perhaps looking into an improved way of inferring the OH concentration and NO2 lifetime… See Figure 1 of Laughner and Cohen https://www.science.org/doi/10.1126/science.aax6832. I would plot NO2 lifetime (calculated from CAMS OH) as a function of CAMS NO2 column data. I am assuming there will be some type of non-linear relationship that can be used to infer the NO2 lifetime when TROPOMI NO2 column data differs substantially from the CAMS column NO2 data. Ideally you'd bin by TROPOMI HCHO which I realize is beyond the scope, but maybe calculating the NO2 lifetime vs. NO2 column relationship by Koppen climate zones could be a quick work-around (which would approximately account for areas with less/more biogenic VOC emissions). This is a long way of saying that if CAMS NO2 has a large mismatch with the TROPOMI NO2 data, your assumed OH may be way off, and there could be an easy way to approximately account for these mismatches.

→ We use here averaged CAMS OH between 950 and 1000 hPa. A study done (Rey-Pommier et al., 2022) with averaging other levels showed that the impact of the thicknesses within which parameters are averaged on total emissions was low. We acknowledge that the method we use here remains basic, and bears errors if CAMS misses or mis-estimates NO_x sources (this would also be seen in a mismatch between CAMS NO_2 and TROPOMI NO_2 as this comment points out). The method suggested by this comment could be an interesting improvement for a future version of the data product, and will be suggested in Section 4.1 "Uncertainties and assessment of results – Model uncertainties" in the revised version of the manuscript.

There is not enough discussion on why biomass burning emissions are not properly captured. It may be worth framing this paper as quantifying fossil-fuel related NOx emissions and purposely screen out areas of known biomass burning NOx emissions, which appear to be particularly uncertain for a variety of reasons (as the authors correctly note).

→ We prefer to keep the title of the title as it is, because there are processes that produce NO_x without involving fossil fuels (e.g. NO_x is emitted in steel recycling using electricity). We also keep it because we might actually capture some biomass burning emissions – but not as properly as fossil emissions, due to factors that are discussed more in details in the revised version of the manuscript. We will however mention this uncertain nature of biomass burning emissions directly in the abstract to avoid any misguidance. We will also cite the following studies on the under-estimation of fire emissions from space: Ramo et al., 2021 (DOI: `10.1073/pnas.2011160118`); Khairoun et al., 2024 (DOI: `10.1016/j.scitotenv.2024.170599`).

In the EDGAR intercomparison, I think small mean bias shown in the "Total" value of Table 3 (i.e…, good agreement) is the product of two offsetting biases: The TROPOMI NO2 operational retrieval is biased low by ~30-50% in polluted areas/cities (Line 468), and NOx emissions are 40% larger at 13:30 local time than the 24-hour average. Therefore I don't dispute your claims in Section 3.3, but I do think that if the TROPOMI retrieval had no bias, then you would be doing an unfair comparison. More clarification should be added. I have added more references and description below.

→ We already acknowledged at the end of Section 3.3 that our comparison between estimates 13:30 LT and daily averages has limitations. We thank the reviewer for the given reference that helped us to detail this point in the revised version of the manuscript. However, a point of clarification is necessary here: in the reference provided in this commentary, mention is made of cities for which TROPOMI is biased low while emissions at 13:30 LT are higher than the daily average. In these cities, emissions are mainly transport emissions. However, in our study, we also estimate emissions from industrial facilities (power plants, cement kilns, etc). Such facilities account for a large part of the global NO_x budget. In addition, they are generally located outside cities, where the TROPOMI bias is lower. Finally, their emissions at 13:30 LT are not necessarily higher than average emissions (this depends very much on the use of the industry in question: some power plants are used for baseload, while others are used primarily to meet peak demand). In conclusion, the effect of the two offsetting biases mentioned by the reviewer is indeed present for cities, but is probably less significant outside cities, in a way that depends very much on the location under consideration. More studies are needed to quantify this effect. In any case, this discussion is detailed in the revised version of the manuscript.

In Section 3.3, it would be interesting to dive a bit deeper into where there is poor agreement between EDGAR and TROPOMI. This would really demonstrate the value of TROPOMI and your method.

→ In the revised version of the manuscript, we develop where EDGAR and TROPOMI estimates disagree the most, and detail the issue of low-income countries that have small diffuse sources or low observation densities.

Detailed comments:

Line 28. A bit more nuance could be useful. You should add something along the lines of "in conjunction with sector- and country- specific NOx/CO2 ratios". There are many examples of NOx emissions dropping rapidly but CO2 not dropping or dropping modestly. I am sure you (the authors) know this but a future reader may not.

→ The text has been changed in the revised version of the article manuscript to account for this comment.

Line 37. The authors are being generous here :-), most bottom-up datasets take 3 years to generate. Unless you know of a emission dataset developed within 1 year, I would default to saying 3 years. This would further demonstrate the utility of your method even if it take several months to process the data.

→ The text has been changed in the revised version of the article manuscript to account for this comment.

Line 82. Which levels of the wind data are used? This is important for study replication.

→ This was not precised in the first version of the manuscript – The first two pressure levels (975 hPa and 1000 hPa) are used for the wind field **w**, hence the calculation of a mean horizontal wind within a layer of about 350 m above the ground. For ground wind $\mathbf{w}_g$ in the topography-correcting term, only the first pressure level (1000 hPa) is used. The text has been changed in the revised version of the article manuscript to precise this.

Line 151. Modify "minor" to "less". I also think you are misrepresenting the Beirle et al. 2019 and de Foy and Schauer 2022 studies a bit as these studies are investigating a relatively small domain over a single season or climatological pattern. A constant NO2 lifetime is not ideal, but a better assumption than if they were global studies. Please correct me if I'm wrong but I don't know of any global study assuming a constant NO2 lifetime. Beirle et al., 2023 uses a latitudinally dependent NO2 lifetime, and I agree your method of using CAMS data is much better. In short, I agree with all your sentiments here, but be careful with some of the nuance.

→ We acknowledge the two studies that are mentioned focus on a smaller domain that justifies the different computation of the lifetime. We therefore changed this text in the revised version of the article manuscript to account for this comment.

Line 202. It'd be best to move discussion in Lines 275 - 278 about wildfires to here. The missing emissions in the Amazonia suggest your method is best for estimating fossil-fuel related sources. Even though Amazonia wildfires take place for only a few months, they should probably show up more distinctly in the annual average than they currently are. Perhaps the days with the largest smoke and NOx emissions are being filtered out as clouds. Another 2-4 sentences are probably needed to discuss these nuances.

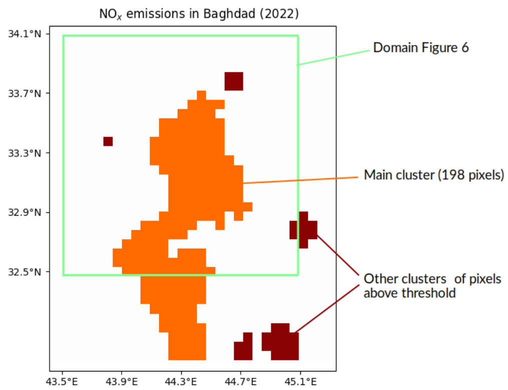
→ The general issue of wildfire emissions is now briefly investigated in the revised version of the manuscript, and the discussion has been moved where indicated by this comment.

Line 203. The sentence "Figure 3…" should be the first sentence of the next paragraph.

→ We prefer to move this sentence in the revised version of the manuscript, but at the beginning of the paragraph where Figure 2 is introduced, as Figure 3 just consists of different zooms on Figure 2.

Line 267. I am confused by how you are counting the number of pixels in a metropolitan area. Using Baghdad as an example, I am counting maybe 30 pixels within the dotted outline in Figure 6, where does the 198 pixels value come from? And can you highlight that 198-pixel "zone" in Figure 6?

→ In the example of Baghdad, the number 198 corresponds to the number of pixels above the threshold of 2 Pmolecules.cm^{-2}.h^{-1}. It does not correspond to the number of pixels in the city core, which is only given as an indication of where emissions are the highest. For such large cities and with this threshold, the cluster generally includes the city core, the corresponding functional urban area, and highways between the city and the main industrial centres nearby. Increasing the threshold to 3 Pmolecules.cm^{-2}.h^{-1} generally makes the distinction between the city core and the rest, as shown in Figure S4 (Supplementary Materials). The Figure below shows the 198 pixels higher than the threshold for Baghdad – The domain is slightly larger than the one in Figure 6.



In the revised version of the manuscript, we emphasize more on what the dotted line stands for, to avoid any confusion. However, we chose not to show explicitly the cluster zone by changing Figure 6 (like above), because we prefer to show the details of emissions on a smaller zone.

Table 1. Typo of Shanghai

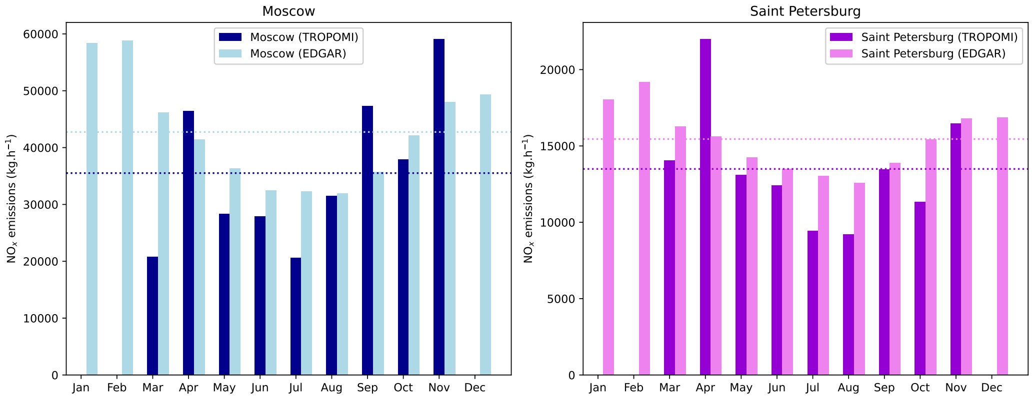
→ The typo has been corrected in the revised version of the manuscript. There was also a typo for Shenz[h]en.

Lines 293 - 325. Thanks for this discussion. There is one policy-relevant question that is still unanswered in this section. From an emissions standpoint, what is the threshold point source emissions rate given a 2 Pmolec-cm-2h-1 threshold? 0.5 tons per hour? Less?

→ Of course the conversion from Pmolecules.cm^{-2}.h^{-1} to ton.h^{-1} depends on the size of the corresponding pixel (~37 kg.h^{-1} at 60°N or 60°S to ~74 kg.h^{-1} at the equator), which is why we prefered working with this unit (Pmolecules.cm^{-2}.h^{-1}). This comment is however relevant, and we give a range of the corresponding threshold in the revised version of the manuscript. It will be added as a comment in the metadata for the user.

Line 355. I wouldn't discount there being a real difference in Russia. How do individual cities in Russia (Moscow, St. Petersburg, etc.) compare against EDGAR?

→ We changed the word discrepancy in the revised version of the manuscript. We also compare below TROPOMI-based emissions to EDGAR for Moscow and Saint Petersburg (domain of ~1.7°×1.7° around the two cities), by summing all pixels with values above 0.2 Pmolecules.cm^{-2}.h^{-1} in the case of TROPOMI-based estimates, as done in Section 3.3:



The horizontal lines represent the annual averages calculated with all daily emissions and excluding NaNs. Note that in January, February and December, no observation was taken above the domain, hence the absence of monthly estimates (for Saint Petersburg in February, a few pixels are observed but they have values below the threshold indicated above). It is also the case for some pixels in the domain in March and November. The order of magnitude is the same for both estimates, and lower emissions in summer (probably due to the lower heating demand) seem replicated. For these cities, the annual emissions therefore do not take into account the winter months, when emissions are particularly high according to the annual profile in EDGAR. The total budget for these emitters might therefore be underestimated. This situation is typical of countries where high emissions occur while the observation density is the lowest. This example will be used in the Supplementary Materials to illustrate biases for large countries with few observations during a part of the year despite having correct agreements for key emitters during the rest of the year.

Line 358. This is consistent with Ahn et al., 2023 (https://iopscience.iop.org/article/10.1088/1748-9326/acbb91) which shows something similar for CO2. I think more detail on this would be interesting and helpful. Which countries in particular show worse agreement? Are they all low-income countries and/or countries with a lot clouds? Maybe a few more of the outlier points can be labeled on Figure 8? I understand why a log-scale is used, but it is a bit deceptive as the Russia bias is probably the largest of all countries. Therefore more discussion in the text is needed.

→ The countries for which TROPOMI estimates are significantly higher than EDGAR are Kyrgyzstan, Uzbekistan, Tajikistan, Zambia, Democratic Republic of Congo, Eswatini, Mozambique, Angola, and Yemen. However, the worst agreements are found for countries where TROPOMI estimates are significantly lower than EDGAR. In Guinea-Bissau, Equatorial Guinea, Togo, Guinea, Gabon, Montenegro, El Salvador, Liberia, Ivory Coast and Myanmar, such differences are higher than an order of magnitude. These countries are countries cumulating low incomes, frequent clouds, and small size. We discuss the potential reasons for such differences in the revised version of the manuscript. We also discuss how results change when no threshold is applied when summing the emissions for countries. Finally, we added more labels in Figure 8.

Line 366-369. Urban NOx emissions at 13:30 are still ~1.4 times larger than the 24-hour average since so many nighttime hours have very low emissions: Please cite and see Figure 4a of Goldberg et al., 2019 which shows an example for New York City, United States: https://acp.copernicus.org/articles/19/1801/2019/acp-19-1801-2019.html I have seen other unpublished studies showing the temporal hourly pattern of GEOS-CF NOx emissions in many global cities look like New York City (and not Seoul). I think you have offsetting biases that are conveniently and approximately cancelling out: The TROPOMI NO2 operational retrieval is biased low by ~30-40% in polluted areas/cities (Line 468), and NOx emissions are 40% larger at 13:30 local time than the 24-hour average. Therefore I don't dispute your claims in Section 3.3, but I do think that if the TROPOMI retrieval had no bias, then you would be doing an unfair comparison. More clarification should be added.

→ We already acknowledged at the end of Section 3.3 that our comparison between estimates 13:30 LT and daily averages has limitations. We thank the reviewer for the given reference that helped us detail this point in the revised version of the manuscript. However, a point of clarification is necessary here: in the reference provided by this commentary, mention is made of cities for which TROPOMI is biased low while emissions at 13:30 LT are higher than the daily average. In these cities, emissions are mainly transport emissions. However, in our study, we also estimate emissions from industrial facilities (power plants, cement kilns, etc). Such facilities account for a large part of the global NO_x budget. In addition, they are generally located outside cities, where the TROPOMI bias is lower. Finally, their emissions at 13:30 LT are not necessarily higher than average emissions (this depends very much on the use of the industry in question: some power plants are used for baseload, while others are used primarily to meet peak demand). In conclusion, the effect of the two offsetting biases mentioned by the reviewer is indeed present for cities, but is probably less significant elsewhere, in a way that depends very much on the location under consideration. More studies are needed to quantify this effect. In any case, this discussion is detailed in the revised version of the manuscript.

Line 390. Thank you for including Portland in Figure 9. First, I am assuming it is Portland Oregon, USA as there is also a Portland, Maine, USA. It is interesting that 84 days vs. 336 days of averaging shows a factor of 2 difference, whereas other cities show less variance by percent. It may be worth commenting that Portland is a relatively small city and cloudy for much of the year, so it's probably "worse case scenario" or "limit" to the type of conditions in which your method works.

→ It is indeed Portland, Oregon. The text has been changed in the revised version of the article manuscript to account for this comment.

Line 425. See prior comment. It is also a function of the size of the city/NOx source too. Large sources may only need one month of data, but smaller sources may need a full year of data.

→ The text has been changed in the revised version of the article manuscript to add the precision indicated by this comment.