The manuscipt titled Quantifying urban, industrial, and background changes in NO2 during the COVID-19 lockdown period based on TROPOMI satellite observations is well-timed and the scientific team very knowledgable on all aspects on this topic. A previously established methodology, on identifying and separating space-based sulphur dioxide observations, is used here as well. However, the manuscript is lacking in numerous respects [see detailed annoted pdf] and does not appear of similar quality as previous works of the scientific team. It is also very long, the supplement enormous, and the methodology needs to be validated/established first, as applicable to tropospheric NO2 observations, before serious statements may be made. I would propose a separation in Part A and Part B, where in Part A, all the methodology/dataset manipulation will be discussed and the first results, for say North America [US & Canada] are presented, followed closely - if not simultaneously - by Part B which discusses the global aspect. I suggest the authors take their time in addressing all comments and deciding how to proceed.

We would like to thank the reviewer for a very thorough and detailed review. We made the corrections suggested by the reviewer. The review also demonstrated that we failed to explain some of the method's features in plain language. We added several paragraphs that provide additional information about the algorithm and its assumptions.

We accept the reviewer's criticism. We focused paper on the new algorithm an on global impact of the lockdown, rather than on comparisons of the obtained results from other publications. In the revised version we provided more explanations and comparisons.

The paper is indeed long, and we had to make it even longer to address reviewers' comments. For example, we had to add 2 more pages to the Data Sets section. We did remove some paragraphs, e.g., related to nightlights. However, even the revised version is not longer than many ACP publications. We prefer not to split it into two papers. The most interesting part is the global overview where the same approach was applied to a large number of areas around the globe.

Note that we also repeated the analysis with a more recent TROPOMI data set and changed the criteria for industrial point sources in the urban areas (used population density directly instead of the correlation coefficients). All of that slightly affected some of the results. We also excluded 2 areas. In one case, the population density data were clearly unreliable, in the other case the noise was too high, probably, due to biomass burning.

p.1 l.18 Is this the total NO2 or the tropospheric NO2? this distinction should already be made at this point.

It is tropospheric NO2 everywhere. Corrected

p. 1 l. 22. In contrast, the decline...

Corrected

p. 1, l. 23 areas

Corrected

p. 1, l. 23 Was the lockdown over in Wuhan as well?

Yes, during a part of the analyzed period. It is in the text, but we do not mention Wuhan in the revised abstract.

p. 1, l. 24 You had point industrial sources within the urban areas studied?

Yes, this is one of the main results of the study that urban and industrial emissions were separated. We guess that the confusion was cause by the term "urban areas". To clarify that, we added that 3 by 4 degree areas were analyzed.

p.2., l 3 Re-phrase, NO2 does not have a long history. Satellite measurements of NO2 have a long history.

Corrected as suggested.

p.2, l. 12. It should be made clear at this point that you will be using the tropospheric VCD. If you do not wish to repeat this phrase, just define it here. Otherwise, it is not clear if you are working with the total VCD.

Corrected as suggested.

p. 3, l. 2 Also on the accuracy of the satellite algorithm? the cloudiness issue? the satellite footprint?

Yes, but this is not the point here. The same satellite data set could produce different decline over an area with, for example, 50 km and 300 km diameter around the city even if there is no issues with satellite data quality. This could be because percent changes over the city center and remote areas could be very different. This is something that we study in this manuscript. We added that to the section where we are talking about the size of the analyzed area here.

p.3, l. 7 Define

Corrected

p.3, l. 14. Your study is hence based on comparing 2020 with the two previous years? at this point, already questions may be raised as to how you dealt with the different meteorological factors that affect tropospheric NO2, such as sunlight, cloudiness, changes in the emissions themselves. Since you allow data with qa > 75, another point is how representative is one year compared to the other year.

We added a sentence that the interannual variability of the three components is discussed later in the text. All statistical uncertainties are very small due to the used method. As for the interannual variability, we investigated it and added more information with two new tables. Previously, this information was in the supplement.

p.3, l. 18. This may be so, however, the TROPOMI algorithm has changed multiple versions between the Phase E1 data [2018], the 2019 data and the 2020 data. I am assuming that further down you demonstrate that there are no such effects in your analysis.

Again, it is just an introduction. But to address your comment, we should point that we compare differences between 2018 and 2019 values and used them as an estimate of TROPOMI NO2 variability that includes all the mentioned factors, differences in the algorithms, number of observations, differences in cloud conditions, etc. This analysis was mostly included in the supplement. In the revised version, it is included in the main text. We also added one more figure that shows that in 2018, 2019, and 2021, on average, TROPOMI NO_2 data where in the range ±2.5%.

p.3, l. 27. Take care to always keep "van Geffen", with small v.

Corrected.

p.3, l.32. OFFL or NRT? provide the algorithm version numbers as well.

The reprocessed (RPRO) and offline mode (OFFL) data were used. The offline version was used for 2019 and 2020 and reprocessed data were used for 2018. We added the version information to the text.

p.4, l.6. State here exactly what you show in Fig. 1 Did you grid the data, for e.g.? one which grid? do not fail to mention what the LT of the observations is. For those not residing in the US, the comment on the "background" values should be expanded, which areas are considered background in the States? last, but not least, there is a known "background issue" with the S5P tropospheric NO2 which produces erroneous tropo NO2 levels. How did you account for that?

The purpose of this figure is just to illustrate tropospheric NO_2 VCD distribution as seen by TROPOMI. We added information about the grid resolution.

p.4, l. 15. Do you mean that you have ERA5 data on the same resolution as S5P, i.e. 3.5x7 and 3.5x5.5km2 after August 2019? I am assuming not, so please rephrase.

It is explained in the next sentence. The wind data were interpolated to the location of each satellite pixel. To avoid such confusion, we replaced "were taken" by "calculated based on".

p.4, l. 17. How did you account for the large spatial difference in the wind data compared to the TROPOMI pixel?

It is explained in the next sentence:” U- and V- (west-east and southnorth, respectively) wind-speed components were then interpolated to the location of the centre of each TROPOMI pixel and to overpass time.” We are not sure what is the question here. It is a common practice to interpolate data to a different grid.

p.4, l. 20. Delete parentheses.

Corrected.

p.4, l.20. That may be so, but the PBL height itself is not constant over the entire domains that you have studied. Are you implying that most of the S5P tropospheric NO2 you analyse resides within the PBL? I am assuming not, since it is well established that the nadir uvvis sensors cannot really sense the PBL. Add a discussion on this matter, and your choice of wind speed height. Between 0 and 1 km seems too low.

The reviewer’s statement that “*the nadir uvvis sensors cannot really sense the PBL*” is wrong. Perhaps the reviewer is thinking of nadir sensors utilizing the thermal-IR? While the sensitivity of a uv-visible nadir instrument to NO2 in PBL is less than in the free troposphere, it is certainly not zero. See Krotkov, N. A., Lamsal, L. N., Celarier, E. A., Swartz, W. H., Marchenko, S. V., Bucsela, E. J., Chan, K. L., Wenig, M., and Zara, M.: The version 3 OMI NO2 standard product, Atmos. Meas. Tech., 10, 3133-3149, https://doi.org/10.5194/amt-10-3133-2017, 2017, Section 3.2, page 3136.

To underscore this, multiple groups have estimated surface NO2 using instruments such as OMI and TROPOMI (Lamsal, L. N., Martin, R. V., van Donkelaar, A., Steinbacher, M., Celarier, E. A., Bucsela, E., Dunlea, E. J., and Pinto, J. P., Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument, J. Geophys. Res., 113, D16308, https://doi.org/10.1029/2007JD009235, 2008.)

The quantity that describes the sensitivity of the NO2 is the air mass factor (or AMF; see Palmer et al., 2001; https://doi.org/10.1029/2000JD900772 for a good description of how AMFs are formulated generally). Here is a plot of clear-sky AMF as a function of height for very low reflectivity (~worst case) and high reflectivity calculated by the authors to help illustrate this. This AMF approach is extremely well documented and is at the heart of the OMI and TROPOMI NO2 (and SO2, HCHO, and others…) data product.

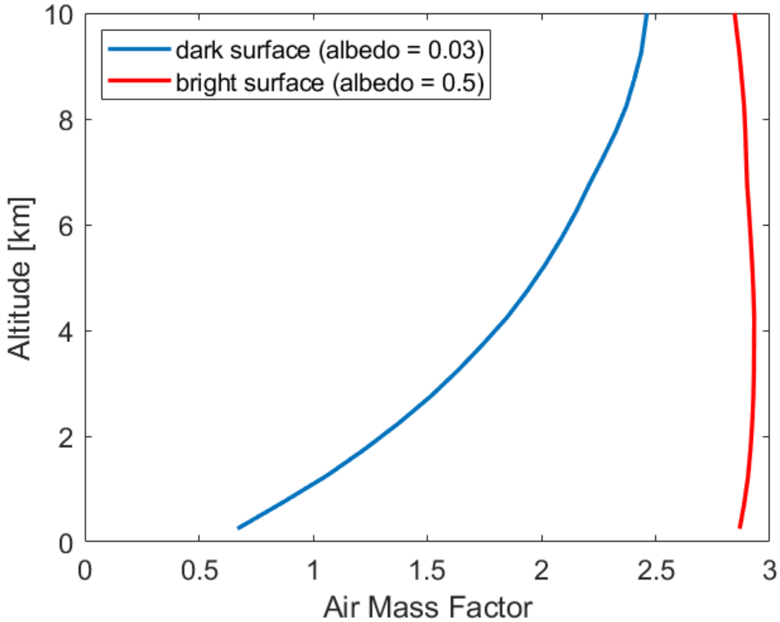

Figure: Sensitivity of a UV-visible nadir sensor to NO2 as a function of altitude (at a wavelength of 440 nm; SZA=50), calculated as in McLinden et al. (2014; 2016) and Fioletov et al. (2016), quantified using an air mass factor. The main point to take away here is the relative change with height, and that even for a less favourable case, there is sensitivity to the surface.

Most of the tropospheric NO_2 indeed resides in the boundary layer (see Levelt et al., 2021). Similar studies use wind data similar to those used in our study. Beirle et al. (2011) averaged from ground up to 500 m, 200 m and 1000 m. Lange et al. (2021) used data from 100 m. Beirle1et al. (2019) used data at 450 m and analyzed wind data at 250 and 730 m for sensitivity studies. We added this information to the text.

p. 4, l. 23. This section requires re-writing, as follows:
1. The title should reflect the sequence the three different auxiliary datasets are presented.
2. The GPW and elevation data are on a much finner spatial analysis than the satellite and the wind data. Provide details on how you dealt with this issue.
3. The TROPOMI data are applicable only for a specific time of day. This is neither the time that people go to work, nor the time that they return. Furthermore, a clear weekday and weekend distinction in tropo NO2 levels has already been shown in literature. How did you use nightlight information on daytime observations?
3. The section describing the emissions is mixed up. First you should state what databases you used, and then how you analysed them.

1. The section was divided into several subsections
2. When lower resolution data were required, they were obtained by averaging the original data within the new grid cells. We added this to the text.
3. Diurnal variations of total column NO2 are rather small (unlike surface NO2 concentrations). We added a sentence about this with some references. Nonetheless, the reviewer is correct, we

do not know anything about nighttime NO2 levels. We also mentioned this. Weekday/weekend differences do not affect the total emissions for the 3-month period analyzed in this study. Even without accounting for this effect in the statistical model, we can explain more that 93% of the variance of the 3-month mean NO2 distribution. But we mentioned these differences in the text. In the revised version, we decided not to use nightlight information. However, it must be noted, that nightlights are one of the main sources of information about urban CO2 emissions and so they should work for our NOx emission estimates as well.

4. It was important to state that we **did not use any emissions data in our estimates**. We needed coordinates of sources only. The actual U.S. emissions data were used **only to validate** our estimates. We changed the text and the section title to make it clear.

p. 5, l. 5. What is the spatiotemporal resolution of these emissions? do you include maps/statistics/information in your supplement? this should be clearly stated here. Have other scientific-based studies been performed using this dataset?

Apparently, there is some confusion here. We **did not use any external emissions data** in the study. Emissions estimates were derived from TROPOMI data. There were only two exceptions: (1) The EPA database was used after the fact in order to verify our TROPOMI-based emissions estimates. (2) A European emissions database was used to obtain the **coordinates** of major emission sources. We modified the text to make this clear.

p. 5, l. 15. What is the former database? the one from 2007-2017? Europe suffered a massive recession during this period, and is definitely not representative to the current 2018-2020 situation. Why did you not use other, well established emission inventories such as the TNO CAMS reg for e.g. for your work? the EPTRT database is quite out of date for some sourced and traffic is not even based on measurements, but statistics. This is a very weak point in your work. How can you separate background, urban, industrial, NO2 emission sources based on this information?

We changed the text to emphasize that no external emission information was used in this study separate background, urban, and industrial NO2 emission sources. The separation was done based on observations only: TROPOMI and wind data, elevation, population density and location of major point sources. This is the key element of this work. **No additional information about emissions or background levels was used.** The emissions database was only used to obtain the **coordinates** of major European point sources.

p. 5, l. 19. This phrase is very poor, very vague, not informative at all. Microsoft Bing can definitely not be cited as such in a scientific paper. How did you detect NO2 sources from Sentinel-2 data? which Sentinel-2 data?

The text does not mention Sentinel 2 data, but satellite imaginary. The sources were detected from TROPOMI (on Sentinel-5p) NO2 data. Point sources appear as “hotspots” on the maps of the NO2 residuals but these alone are often insufficient to locate the source to better than a few km. The satellite true-colour **images** were used to determine the precise locations of the sources. Google maps are widely used in similar studies (e.g., McLinden et al., 2016; Beirle, et al., 2021).

The problem is that often the imagery available through Google Maps is not up to date and do not show recently build sources. Microsoft Bing maps are similar to Google maps, but in some cases, it has newer images. For some factories, they also have information about the source owners and type of the business that helps to determine the type of the emission source. This combination of imagery and ancillary information is very useful for pinning down source locations.   Sentinel 2 provides daily images of the Earth surface. They are up to date, but have lower resolution. Once locations of such addition sources are established, they are added to the fitting procedure and the residuals were examined again to check if there is any improvement (i.e., reduction in the residuals over the hotspot). If yes, the source was added to the point source list. We added the links to the used satellite image resources on the Web.

p. 5, l. 28. You have to make a comment also on the diurnal variability of NO2. SO2 is an entirely different gas, its sources are constant during the day, the weekend, NO2 is nothing but. This method may be valid for large SO2 sources, but this does not necessarily mean that by default it can be used on NO2 measurements taken at 13:00 LT, i.e. at the lowest level of the NO2 diurnal variability, which means that the background NO2 levels are far below the detection limit of the TROPOMI algorithm.

There are numerous previous studies of NO2 emission from cities estimated from satellite data. The reviewer is mixing up surface NO2 concentrations, that have a strong diurnal cycle and tropospheric NO2 VCD. The latter have much smaller (if any) diurnal variations and measurements taken at 13:00 LT by no means represent "the lowest level of the NO2 diurnal variability". Please see figure below that shows diurnal variations of tropospheric NO2 over Korea in May-June. It is from Chong, H., Lee, H., Koo, J.H., Kim, J., Jeong, U., Kim, W., Kim, S.W., Herman, J.R., Abuhassan, N.K., Ahn, J.Y., Park, J.H., Kim, S.K., Moon, K.J., Choi, W.J. and Park, S.S. (2018). Regional Characteristics of NO2 Column Densities from Pandora Observations during the MAPS-Seoul Campaign. Aerosol Air Qual. Res. 18: 2207-2219. https://doi.org/10.4209/aaqr.2017.09.0341

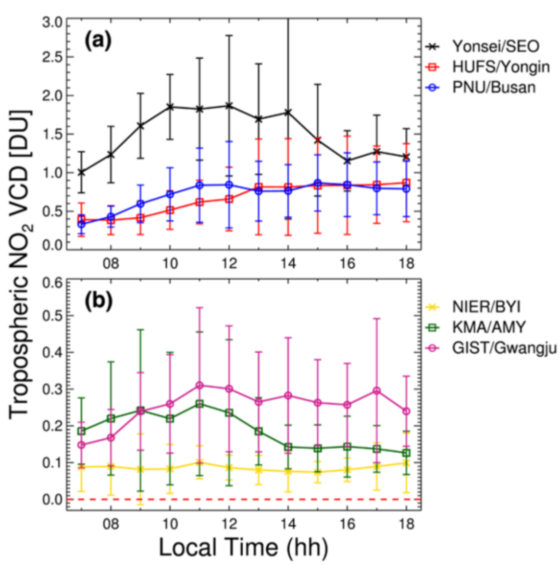

Fig. 9. Diurnal variations in tropospheric NO_2 at (a) Yonsei/SEO, HUFS/Yongin, and PNU/Busan, and (b) GIST/Gwangju, NIER/BYI and KMA/AMY, based on hourly averaged data. The error bars represent the standard deviations; the red broken line in (b) indicates 0 DU.

There are other studies with similar results (see figure below): Herman, J., Cede, A., Spinei, E., Mount, G., Tzortziou, M., and Abuhassan, N. (2009), NO2 column amounts from ground-based Pandora and MFDOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation, J. Geophys. Res., 114, D13307, doi:10.1029/2009JD011848.

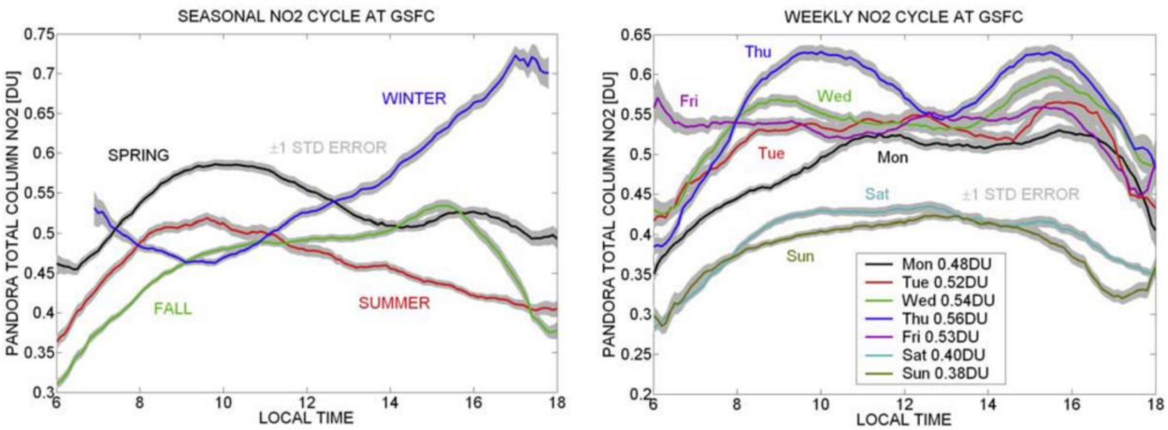

Figure 18. (left) Average seasonal time dependence of $C(NO_2)$ during the day. (right) Annual average for day of the week diurnal behavior of $C(NO_2)$ at GSFC. Winter (December, January, February), spring (March, April, May), summer (June, July, August), and Fall (September, October, November). The gray area around each line shows the mean value ± the standard error of the mean.

The fact that weekend urban emissions in weekends are lower that on weekdays does not mean that we cannot estimate **total** weekly emissions. This algorithm was successfully used to monitor SO2 emissions from volcanic degassing, where emission variations from day to day can be much large than 20-30% differences between workdays and weekend emissions. See for example Fig. 9 of Fioletov et al., (2016).

Studies of weekend/weekday differences (e.g., Lange et al., (2021)) are based on several years of data. Twelve weekends during the period in 2020 analyzed in this study are not enough for such estimates.

Finally, the algorithm itself partially accounts for diurnal variations. Recall, that it is based on fitting the plume as a function of distance from the source. If emissions at the source were variable, it would appear as deviations rom the EMG fitting function shape and the algorithm would produce "average" emission value.

Of course, the reviewer is correct in a sense that there is a systematic difference in daytime and nighttime urban emissions, and it is hard to study such difference from a single satellite overpass per day.

We added a few sentences about NO2 variability and stressed that we are talking about daytime emissions.

p. 6, l. 5. Why such a large area? if you were planning on showing only for e.g. the east coast of the US, India and China, I would say, fine, but you claim at far smaller sources, plus the separation of courses. This begs the question.

The question is not clear. There could be multiple industrial sources within the analyzed area and emissions from each of them are estimated. The larger the area, the less accurate assumptions about a linear gradient of background NO2 and constant emissions per capita are. The algorithm is based on fitting of plumes. For a lifetime of 3.3 hours and wind speed of 30 km, the NO2 VCD in the plume declined *e* fold on a 100 km distance. So, the size of fitting area should be of order of at least 100 km. Moreover, the area should be large enough to avoid correlation between the elevation and population density. For all these reasons, a 3 x 4 degree area was selected.

p.6., l.24. What is Supplement A? there are numerous files in the supplement but no Supplement A.

There is only one text file on the Supplement that contain several sub-sections. Apparently, some of the ACP rules have changed recently. We revised the Supplement and removed additional files.

p. 7, l. 9. A better agreement between what and what?

It gives a better agreement of the fitting results with the satellite data. We added “the fitting results”.

p. 7, l. 18. This can definitely not be considered valid for many parts of the world. How do you explain this choice?

We need to derive such emissions per capita estimates first. The algorithm produces the mean value for the emissions per capita for the area. If the assumption of uniform emissions per capita is not valid, it would appear on the maps of residuals: the areas of emissions per capita above the average for the whole area would show positive values, while area of emissions per capita below the average would have negative values. We examined the maps of residuals for all 261 and this is clearly not a common problem. In case of the area near the border between South and North Korea and in a few other cases, the fitting area was manually adjusted to be located within one country. We added some comments to the text.

You also should be aware of the context. All previous studies of urban NO2 emissions analyze a city as a point source, without consideration even the population distribution within the city.

p. 7, l. 19. So, you multiply every 0.2x0.2 pixel with the same number within the 3x4 area you analyze? does this make much sense, if you want to claim that your method "sees" point NO2 sources?

This sentence does not talk about industrial point sources that we can “see”.

This paragraph discusses Ω_p, the contribution of the urban emissions based on the population density. We do not multiply the 0.2x0.2 pixel with the same number. We consider each grid point of the 0.2° by 0.2° grid as a potential emission source and calculate the contribution of emissions from that grid to each TROPOMI pixel. That contribution is based on the position of the source relative to the given TROPOMI pixel and the wind speed. So, for each TROPOMI pixel, we have a contribution from 336 (16 x 21) “sources”. The absolute value of such contribution depends on emission strength of each “source”. It is unknown, but we assume it depend on the total population at that grid point. Then we are finding that unknown emission strength parameter from the best fit of our assumed emissions to all TROPOMI pixels collected during the analyzed 3-month period within the analyzed 3 x 4 degree area.

We added a paragraph that discussed the algorithm in more straightforward terms

p. 7, l. 27. Not so for Milan and the Po Valley.

We are not sure, what is this comment related to. There are no significant NO2 industrial point sources in the Milan area. If the reviewer is talking about the correlation between elevation and population density for the Milan area, it is not high, only -0.35. The contribution of individual components to the NO2 distribution in the Milan area is in Figure 11. We can see a very clear separation of the background and urban components.

p. 7, l. 32. So, you chose this grid to account for cities in valleys surrounded by high mountains? how many such cities do you analyze, in the end?

We analyzed 261 areas and each of them was 3 by 4 degrees. We added a paragraph that explains the area size selection.

p. 10, l. 5 Check this statement. At 40 North I would expect a larger area in longitude.

We replaced 40 with 42.

p. 10, l. 18. Delete parentheses.

Corrected.

p. 10, l. 18 Please add a reference for this choice. There are other established works that suggest a much smaller percentage, 3% and not 30%. Explain your choice.

The NOx/NO2 ratio was estimated by Beirle, S., Borger, C., Dörner, S., Eskes, H., Kumar, V., de Laat, A., and Wagner, T.: Catalog of NOx emissions from point sources as derived from the divergence of the NO2 flux for TROPOMI, Earth Syst. Sci. Data, 13, 2995–3012, https://doi.org/10.5194/essd-13-2995-2021, 2021.

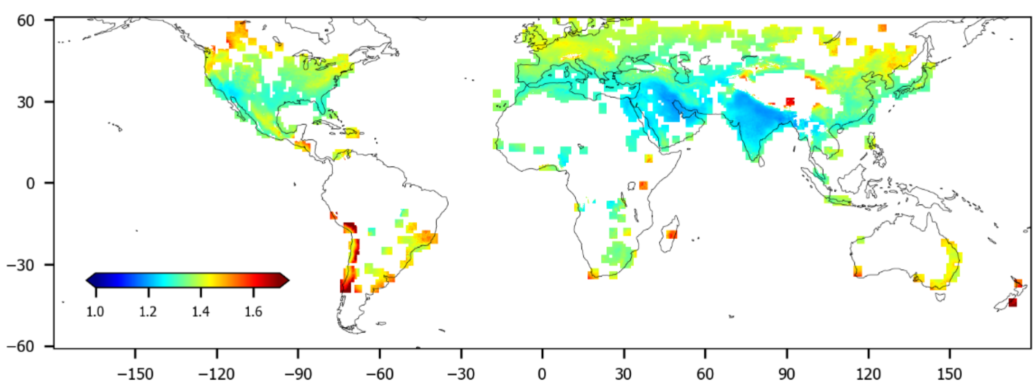

Figure 2. Effective NO_x/NO_2 ratio, i.e., mean tropospheric NO_x (as derived by assuming photostationary state according to Eq. 2 for each TROPOMI pixel) divided by the mean NO_2 column density for 2018–2019. Note that only cloud-free observations with SZA $< 65°$ are considered; thus wintertime measurements at mid- to high latitudes are skipped, and the expected latitudinal dependency of the NO_x/NO_2 ratio is suppressed. The spatial variability is a consequence of the dependency of the photostationary state on actinic flux, ozone concentration, and temperature (Eq. 2).

As you can see from their Fig.2, the ratio is between 1.2 and 1.6 practically everywhere in the world with the largest majority between 1.3 and 1.5. In our study, we used the value of 1.0/0.71=1.4. The reference to Beirle, et al., 2021 and other studies was added.

p. 10, l. 18 Write a small introductory paragraph, three lines, to state what you will show in this section. You start with the cities, you give results, and at the end you add another city which is a different type of case study. Not very reader-friendly.

We added an introductory sentence.

p. 10, l. 30. Houston.

Corrected

p. 10, l. 30. How is this explained? an oversight from EPA? did you contact them about it?

The used EPA NEI CEMS database contains emissions that are **directly measured** at facilities stacks with hourly resolution by sensors. These are the most accurate emissions data. Since we used reported emissions data for validation our own emissions estimates, we opted to use only the most reliable emissions data that are based on actual stack measurements. In the revised version we added less reliable annual emissions from EPA eGRID inventory, for sites that are not in the CEMS database.

p. 11, l. 5 1. Does this value agree with the other studies you have already referenced in your paper?

We added such information

2. What values was this percentage calculated on? seasonal kt? per pixel? per total? on the 3x4 grid? over Boston only?

It was discussed in section 3. We added a plain language summary of section 3 at the end of that section. We also added comparisons with other studies.

3. Is there an std on this percentage? surely something could be estimated for the wind data, the S5P tropo NO_2 data... please discuss.

The uncertainties were discussed in the supplement. The statistical uncertainties are very small, about 1%-2% and it is hard to make them visible on the plot. The real problem is interannual variability due to factors such as sampling, meteorology, etc. In the revised version, we moved the uncertainty estimates to the main text and added Tables 1 and 2. That includes both random uncertainties and interannual variability.

p. 11, l. 17. Since you provide a numerical result for Boston, why did you now provide something for Atlanta already?

We added some numbers and comparisons. The numerical values for all areas are also shown in Fig. 8.

p. 11, l. 8. ... busiest airport with....

Corrected.

p.11, l.19. Their emissions are....

Corrected.

p.11, l.24. Since you provide a numerical result for Boston, why did you now provide something for Pittsburgh already?

The numerical values for all areas are shown in Fig. 8. We added some numbers to the text as well.

p. 11, l.27. Known from where? if they are known, why are they missing from the EPA database? do other emission inventories, commonly used for e.g. in the GEOS-CHEM model, include these sources?

Please see the comment to p. 10, l. 30. above.

p.11, l.31. Not a proper way to reference the supplementary material. Check with the journal instructions.

Corrected.

p. 12, l. 4. What are the numerical findings here? At what distance are the 2000m high mountains away from Seattle? how spread out is Seattle to begin with? ditto for Mineapolis.

The purpose of this figure is to illustrate the NO2 distribution in a flat area and in the area with high mountains. In the case of Seattle area, NO2 distribution is dominated by the local topography. The numerical values are given in the next section. We added the elevation and population density maps for the two areas and moved this figure up.

p. 12, l. 8. How is this explained, by physics but also knowledge of the location? For someone not aware of the geo-physical parameters of Montreal, nor the population spread, this statement reads a bit strangely.

This is just a statement based on estimates from TROPOMI NO2 data. We added some additional comments. We just commented on what was shown in Fig 2 d, e, and f.

p. 12, l. 14. I strongly suggest that you clearly discuss that this whole method is based on one time of measurement per day and then the rest of the times are assumed on a known diurnal variability of NO2. The power plants may emit the same every day, but every day is not photochemically the same as the previous one.

In this study, we analyze the observed NO2 distribution in and around major urban areas. Such analysis can be done only on a relatively long time intervals such as a season. So, the results represent mean characteristics of the NO2 distribution. The fact that emissions may be different from day to day and/or that one day is photochemically different from another simply means that we cannot get daily emissions with this method because the uncertainties are very large. In order to make the uncertainties smaller, we had to combine multiple days. In this study, we used a three-month interval. The reviewer is right that the results represent NO2 distribution at the time of satellite overpass and all other characteristics are derived from such one-pass measurements. We added some discussion about this.

p. 12, l. 25. Indeed, but it mainly shows that the background levels are far greater than either. Where is this background NO2 coming from? this finding clearly goes against all the global efforts in cleaning power plant emissions, cleaning car emissions. etc. Your paper in this point should address this issue extensively.

We have some discussion about the background component at the end of the manuscript. Our goal here is to demonstrate the magnitude and shape of the background component. Keep in mind, that previous studies simply removed that mean background component from the analysis and focused on the plumes only. We added some text to this section.

p.12, l. 27. Please add information on the dates, measures, etc. for COVID in 2020 over US and Canada. This will help the reader understand your results. Did people really stay indoors? what was the original TROPOMI variability, in molec/cm2 over the cities for the three times you studied? what was the meteorology like? you cannot simply provide the statistical results without links to the actual situation.

Same is needed for the other areas you study.

Since we used a 3-month period, the exact date of the lockdown is not very important. Practically all countries had some restrictive measures during that period. We added information about lockdown periods and added a figure of the mobility changes based on Google data. For the analyzed period, Google mobility data were below the baseline level in twelve analyzed regions. China was an exception since the lockdown period there was earlier and Google data were not available there.

The methodology was described in section 3. For example, for the urban component, all TROPOMI measurements for each 3-moth intervals were reduced to just one number, that can be expressed as total emissions per 3^{o} by 4^{o} area or as emissions per capita. We added some additional plain-language explanation.

Information about measurement uncertainties and their impact on the derived characteristics is included in the revised manuscript.

p. 12, l. 29. This is not what is shown in the figures, or the figures are erroneously labeled. I am guessing the latter. In any case, you need to carefully check what you show and what you describe.

We changed the labels. We added comments on how the values shown in Fig, 8 are linked to the parameters discussed in Section 3.

p.13, l. 5. Can you use this result to suggest where this background is coming from?

We have some discussion about of the origin of the background component at the last section. It is more appropriate to discuss it after the results from all regions are shown.

p. 13, l. 17. Check this.

Corrected.

p. 13, l. 20. ...emissions...

Corrected

p.13, l. 22. What is supplement B?

The supplement was edited. There are no sections in the new version of the Supplement.

p. 13, l. 27. You definitely need to discuss this a lot further. While it is true that the industrial sector shut down entirely in many Chinese provinces for a few weeks, this was definitively not the case for other parts of the world. Surely this information is known for the US and Europe, You need to add a full discussion on this point.

You also need a full discussion on how your US results compare to other studies which you already reference in the intro.

We would like to remind the reviewer that we estimate emissions only for industrial point sources located within the analyzed 3 by 4 areas around large cities, not in the entire US. The goas here is to separate urban and industrial sources. And we estimate these emissions. To address your concern, we added comparison with the latest 2020 EPA emissions inventory.

p. 13, l. 32. Add a Thiel-Sien line here, and all the associated statistical measures. Which reported emissions are those? EPA only? state again clearly.

We added more information to the plot and the regression line. A Thiel-Sien line is not necessary here since the residuals are Gaussian.

p. 14, l. 1. Include more statistics here.

We added a section 4.4 with Table 1 that contains additional information.

p. 14, l. 21. A number of studies are already published on the effects of the COVID-related restrictions, including per country analyses, Italy, the UK, Spain, etc. You have referenced some of them already but you need to clearly add how your results compare to other studies. A quick search from scopus resulted in:

We added some discussion here. The selection of suggested paper is very strange: some of them are comprehensive studies of the COVID impact, while the other are limited to results of TROPOMI NO2 data mapping.

We would like to remind the reviewer that some of the mentioned papers just **state** that it is important to take the background NO2 into account, while we **estimate** that background component and its changes.

p. 12, l. 3. Is this expected? did the European industry really shut down? all of it? some of it? which? add a full discussion on this statement.

We added some discussion with numbers here.

p. 15, l. 5. It cannot be assumed that the readers are aware of the Manchester-Birmingham area, neither its population, the spread of said population, nor the industrial area that surrounds it. More detail in discussion is needed here.

Same comment applies to all the chosen case studies of this section.

We added some additional information. However, we are not sure that additional information about “the spread of said population” is really needed since the column ‘e’ itself is the population density map convoluted with EMG functions driven by the wind field. We added a reminder about that.

p.15, l. 34. From this paragraph, one understand that the reduction in emissions is environmental policy-related and not COVID-related. Rephrase and explain your meaning.

We are just saying that in power generations from coal-burning plants was reduced by 60% in 2020 compared to 2019.

p.16, l. 12. Not many European cities can claim a population over 6 mil in the metropolitan area. How do you justify this choice for China?

Are you now claiming that your method is best applicable to larger sources? if so, what are the scientific criteria? you can add a table discussing this feature, since you found that TROPOMI cannot sense changes over the "smaller" hotspots.

It appears, there is some confusion here. We considered all cities with population greater than 1 million. However, there are too many of them in China (compared to the other regions). Since the study is not focused on China, we included only 21 Chinese cities with population over 6 million in our analysis. In this case, the number of analyzed cities in China is comparable to the number of cities in other regions.

We produce emission estimates for all 261 areas analyzed in the study. The paragraph discusses why several cities with population over 1 million are **not** included in our analysis.

We made some changes to make this clear.

p. 16, l. 18. Update this reference and wording according to journal standards.

We excluded that file from the Supplement

p.16, l. 32. Was the winter of 2020 colder than the average of 2018-2019? you have to comment on this point in more detail, otherwise your method cannot be seen as valid over such regions.

Figure 13 does not discuss 2020 data. It is clearly stated in the figure caption. The sentence is related to the fact that emissions per capita over Russia are higher than over other regions of the globe. We added a few words to clarify that.

p. 17, l. 11. Because.... ?

Because they emissions were very different from those from other countries in the region. Corrected

p. 18, l. 13. This comment applies to the entire section above. You have to comment on other studies on the same topic and how their results comply/agree/disagree with these results. This is a a pivotal point in any scientific presentation.

We added some comparisons and discussions. We also had a comparison with a similar study by Lange et al., 2021 that was previously in the supplement.

p. 18, l.29. The dates where these activities stopped due to COVID differed greatly between your sites. How can you comment on this fact?

There are different approaches to the COVID-19 lockdown impact analysis. We are trying to answer the following specific question: “How were NO2 emissions and levels during the period

March 16 -June 15, 2020 different from the 2018-2019 baseline for the same period in urban areas around the world?"

As was already stated, that **all** analyzed regions (except China) were affected by some form of lockdown and decline in mobility during the analyzed 3-month period. Even if the lockdown was not formally introduced. We added a plot that shown changes in mobility based on Google data to illustrate that (new Fig. 3).

Clearly, a detail analysis of each of the 261 analyzed areas is not possible in such study.

p. 19, l. 15. This finding is rather dubious. How did you deal with your daytime obs compared to the nighttime lights? too many assumptions are included in this comparison. I would suggest you remove this entirely and present a different, in detail study using night lights, to show unequivocally their usefulness in such comparisons.

It appears that the reviewer is not aware that some global CO2 emission inventories are based on satellite nightlight data. The nightlight dataset is just a proxy for urban emissions or population density. In any case, we removed this paragraph to make the paper shorter.

p. 21, l. 8. You used many more datasets which should be explicitly referenced here. Which TROPOMI product did you use? state clearly here as well.

We added references to the other data sets. The version of the TROPOMI product is added to the Data section.

p. 39, l.2. This description does not reflect what the plots show. The upper left plot shows TROPOMI NO2 with incomprehensible units and not std bars, to boot. The middle and bottom left plots are different altogether, per capita and per kt/annum. Please correct this, and the text as well.

We explained that the random uncertainty, which is typically shown in error bars, is very small for the background and urban components (within a couple of percent). There is no reason to show them on the plots. They are included in Table 1. The real issue is the natural variability. It is also shown in Table 1 and discussed in section 4.3. We also added the interannual variability limits as the grey dashed lines to the figures.

We added some plain language explanation that the panels show **the same type of quantities** that were derived from the regression Equation Eq1.

p. 45, l. 6. Can you add the std of the mean value for the sites you chose? the correlation is obvious, and remarkable of course, but statistically you should also add this information.
The analysis of the figure needs improving and the title on the y-axis written properly, explicitly, etc. One has to be able to understand from the graph what is shown, as first information, without needing to read the text.

The error bars are added, and the axis labels are expanded.