Response to Referee #2

We would like to thank reviewer #2 for taking the time to review this manuscript and for providing valuable, constructive feedback and corresponding suggestions that helped us to further improve the manuscript.

In this authors' comment, all the points raised by the reviewer are copied here one by one and shown in blue color, along with the corresponding reply from the authors in black.

The manuscript presents an interesting new method for the determination of spatially resolved emission maps around megacities based on TROPOMI NO2 observations, wind fields from meteorological models, and machine learning techniques. The study matches the scope of AMT. Before publication, however, major additions/extensions are needed.

The paper presents results of the new approach exemplarily for Riyadh and Madrid and states that the method "works properly and is reliable" (line 283).

However, the resulting emission maps reveal strong artefacts which are not at all mentioned in the paper.

A critical evaluation/discussion of shortcomings, artefacts, problems, and uncertainties of the proposed approach is missing.

Major concerns:

1. The presented results reveal several artefacts:
   a. several pixels of quite high emissions over regions without obvious NOx sources, e.g. at 25.05°N, 46.45°E and 25.25°N, 46.45°E (Figs. 1d, 3a).
      There is neither a significant NOx emission over this area reported in Beirle et al., 2019, nor is there a local enhancement in the NO2 column (Fig. A4).

Figure R1 presents terrain map and spatial distribution of NO_2 columns on a fine resolution. Several highways cross the region at 25.05°N, 46.45°E and 25.25°N, 46.45°E (in yellow) and there are some residential areas scattered in between, especially on the west of Road 535, like Malham city. There are several ready-mix concrete companies in the blue rectangle region on the east of Road 535. Mean local enhancements can be seen from the TROPOMI observations, though they are weaker than those near to the city center. However, the normalized temporal variability in these regions is far larger than that in the city center. This is maybe what is accounting for the enhanced emissions in this region, following Figure R2.

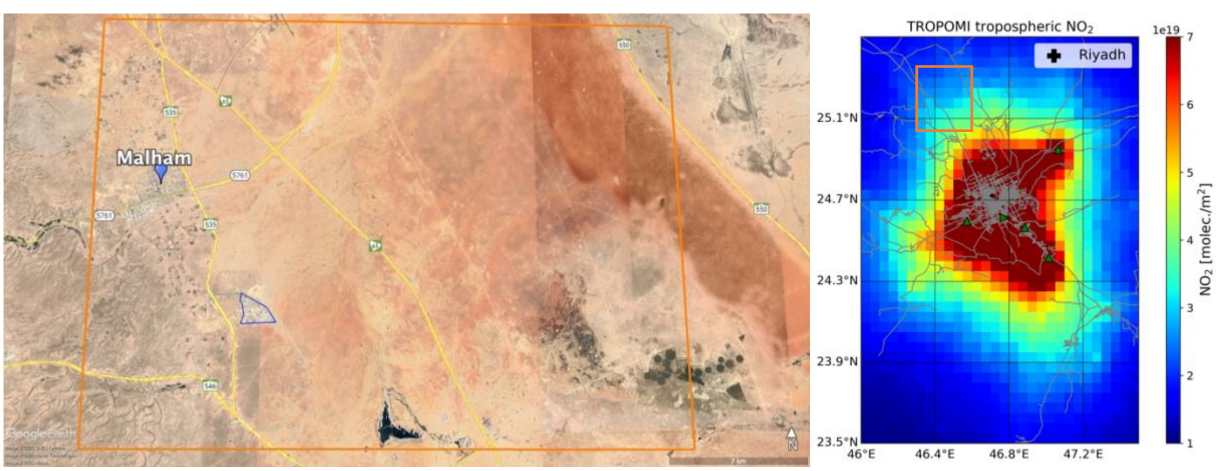


Figure R1: Terrain map (left) and TROPOMI tropospheric NO_2 column on a regular latitude-longitude grid with 0.05° spacing (right) in Riyadh. Red rectangles represent the area between 25.05°N - 25.25°N and 46.30°E - 46.60°E.

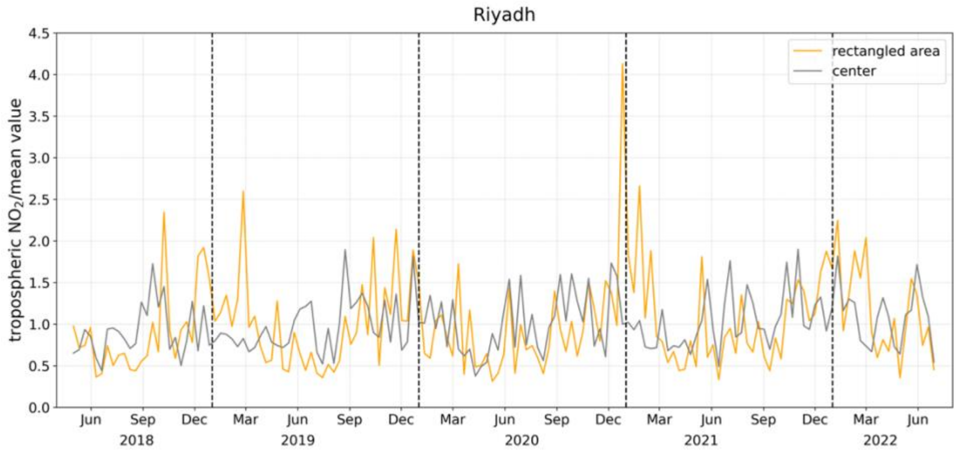
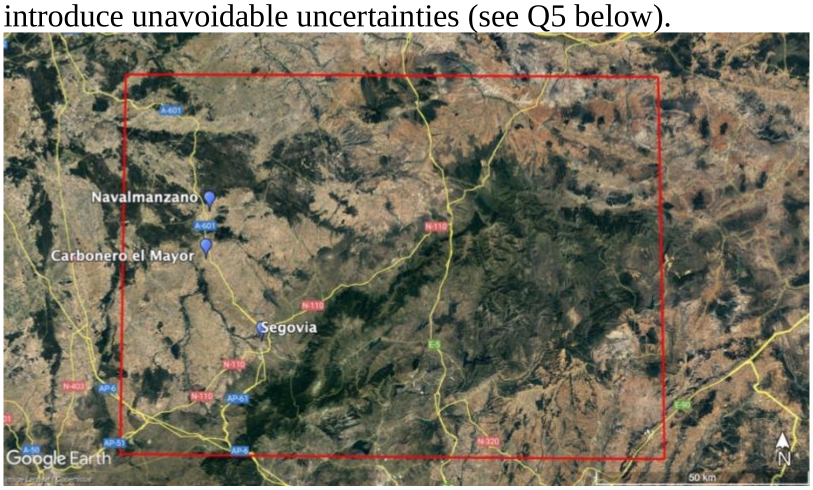


Figure R2: Time series of variations (tropospheric NO_2/mean value) for the area in center (24.5°N - 24.9°N and 46.6°E - 47.0°E) and for the rectangle area in Figure R1 (25.05°N - 25.25°N and 46.30°E - 46.60°E).

b. a large extended area of positive emissions north of Madrid (>40.7°N), where CAMS emissions are close to zero.
   While values of individual pixels still look relatively low in the color coded image, the integrated emissions >40.7°N are still considerable, and I do not think that these emissions are real.

When looking at the TROPOMI observations at a fine spatial resolution, there are considerable local enhancements north of Madrid (>40.7°N) (Figure R3). The region is crisscrossed with highways and roads, connecting small cities or villages, like Segovia, Carbonero el Mayor, Navalmanzano along the highway A-601. The variations in two regions are different (Figure R4) in different seasons, which indicates different emission sources (e.g., biomass burning, construction work, fertilization, etc.). City center region shows some bias low in May-September and bias high in December-March. Thus, some positive emission north of Madrid can be expected. But the referee is correct that there are errors existing. We applied a simple empirical model for calculating enhanced NO_2. It is determined by angle α which is an empirical value, and average decay time τ whose seasonal and spatial variability is not considered. These two parameters introduce unavoidable uncertainties (see Q5 below).



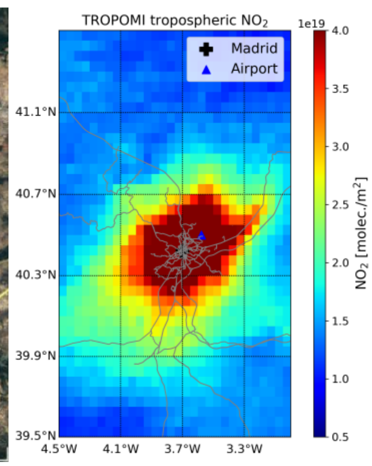


Figure R3: similar to Figure R1 but for Madrid. Orange rectangle represents the area in north of Madrid (>40.7°N)

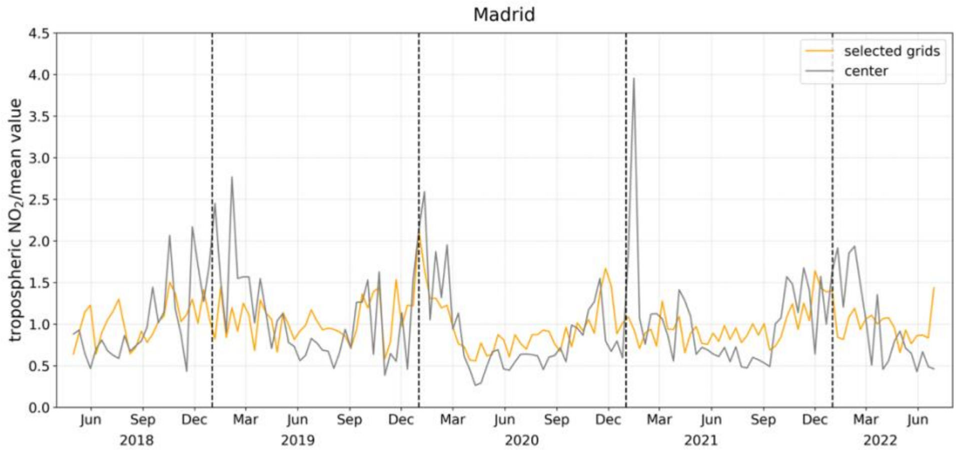


Figure R4: Time series of variations (tropospheric NO_2/mean value) for the area in center 40.3°N - 40.5°N and 3.9°W – 3.5°W) and for the grids with dominant emissions (>1.0E23 molec./s) in north of Madrid (>40.7°N).

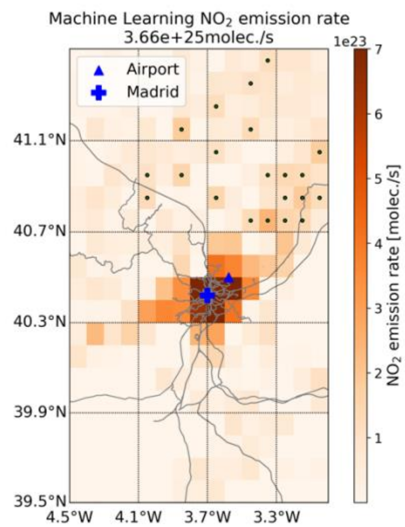


Figure R5: estimated emissions in Madrid. The grids with dots represent the emission rate > 1E23 molec./s in north of Madrid (>40.7°N).

c. generally "checkerboard-like" structures in the maps for data subsets.
   This indicates a problem with the method that involves solving a linear equation iteratively. It seems that initial deviations are overcompensated in the next neighbor, than undercompensated in the 2nd next neighbor, and so on, indicating an instable system with oscillating values in the solution. I think this effect is a known problem for inverse problems, and the authors might check whether they find standard procedures for avoiding or supressing these oscillations.
   In any case, the authors have to clearly describe the artefacts and discuss possible reasons.

The Gradient Descent (GD) approach tends to obtain the minimal biased (loss function) between true values (satellite observations) and modeled values (wind-assigned anomalies). The loss function is decreasing during the iteration process (see Figure R6). In our study, up to 80000 iterations are performed until reaching the minimal bias. Thus, there are no oscillating values in the solution.

The uncertainties are discussed in an additional subsection in the manuscript.

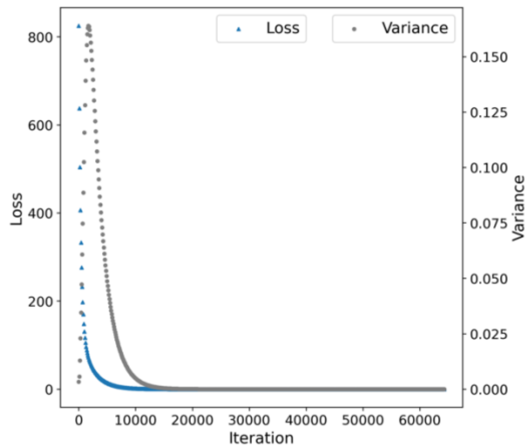


Figure R6: Loss and variance as a function of number of iterations.

As already demonstrated, the temporal variability contributes significantly to the emissions. What would be most interesting would be for follow-up work relating to this (Wang et al., 2021). Is it being caused by fires, but new urbanization, or something else? However, thanks to the deeper review called for by the referee, we now are more confident that these results actually are emissions, and not just noise.

The sentences have been added to the conclusion in the updated manuscript:

"We assume that the checkerboard-like structures indicate that the inversion attempts to resolve fine structure which is poorly constrained by the observation. When we converge to a stable solution with minimal bias, we are confident that spatially averaged retrieved emissions are realistic."

Wang, S., Cohen, J. B., Deng, W., Qin, K., & Guo, J. (2021). Using a new top-down constrained emissions inventory to attribute the previously unknown source of extreme aerosol loadings observed annually in the Monsoon Asia free troposphere. Earth's Future, 9, e2021EF002167. https://doi.org/10.1029/2021EF002167.

2. The authors should be more careful about the usage of the terms "NO2" and "NOx". Emissions are sometimes referred to as NO2 and sometimes as NOx in the manuscript. Clarify the issue of NOx = NO + NO2 in the beginning (emissions should refer to NOx, while TROPOMI only measures NO2). Specify how you account for the missing NO in the NOx budget. Note that there are more "oxides of nitrogen" (line 36) such as NO3, N2O5 or N2O, which are not included in NOx.

Thank the referee for pointing this mistake which might mislead readers. We have clarified the issue of NOx = NO + NO_2 in the abstract and in the introduction. The misleading phrase of "oxides of nitrogen" is removed as well. We specify the emission of NO_2, and the use of NOx is corrected to NO_2.

3. In the introduction the authors give a quite high range for the lifetime of NOx of 1-12 hours (should be "tropospheric" rather than "atmospheric" lifetimes in line 43).

Thanks. We changed the "atmospheric" to "tropospheric".

However, later they just use one fixed value, ignoring probable seasonal and spatial variability of the lifetime. This simplification has to be stated clearly and the impact on the resulting emission maps should at least be investigated with some simple case studies.

The referee is right that the probable seasonal and spatial variability of the lifetime exist, but it is not considered in the study. To make it clear, this information is added to Section 2.2:

> "v is the wind speed from ERA5 and τ is the lifetime/decay time for NO_2. For simplification, seasonal and spatial variability of lifetime is not considered, and empirical values based on Beirle et al. (2019, 2011), i.e., fixed values of 4 hours for Riyadh and 7 hours for Madrid, are used in this study."

The choices of τ are adopted from the studies of Beirle et al. (2011, 2019). We discussed the uncertainty of the choices of τ as the α in the additional section 4 in the updated manuscript.

> "The angle (α) of the emission cone is an empirical value, so as the lifetime/decay time (τ) for NO_2. They can introduce uncertainties and thus, different values for α and τ are used to investigate their impacts on emissions. The spatial patterns of the estimates with using different α or τ are quite similar. The absolute values of emission rate increase with the increasing α (see Figure 7-left). A change of 10º in α introduces a difference of less than 3.2%. A decrease of 1.5% is observed when using $\alpha = 50º$, and an increase of 1.4% is observed for $\alpha = 70º$, as compared to $\alpha = 60º$. The increasing values of τ result in lower estimates (see Figure 7-right). With respect to the result obtained with $\tau = 4h$, the estimate increases by ~42% for $\tau = 3h$, and it decreases by ~20% for $\tau = 5h$.

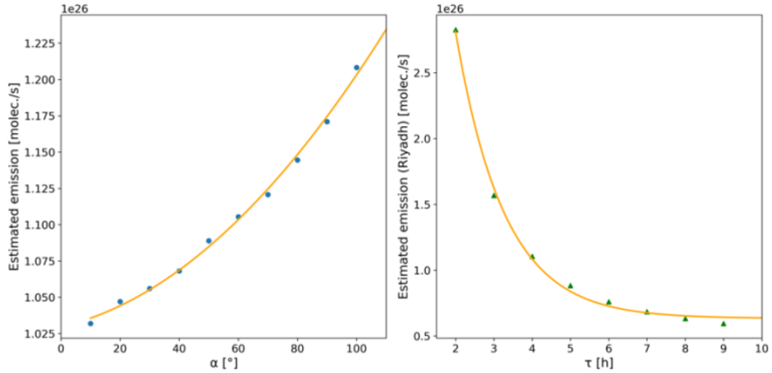


Figure 7: Estimated emissions under different cone angle α (left) and NO_2 lifetime τ (right) based on TROPOMI data in Riyadh in 2019."

4. The study makes several simplifications such as constant lifetime, constant wind field, no consideration of seasonal effects (which might correlate with wind direction and thus would directly affect the wind-assigned anomaly). A discussion of the impact of these simplifications, and in general an error discussion is missing.

The referee is right that we simplified the method by using constant lifetime (τ) and opening angle (α). For wind field, the time-resolved variation has been considered based on the ERA5 wind which has a temporal resolution of 1h. Only the spatial variation of the wind field across the scene is ignored for limiting the computational effort and this impact is expected to be small (1.9% in Riyadh and -1.3% in Madrid). We have added a section (Section 4) about the uncertainty analysis as the referee suggested. The constant wind field in spatial distribution is discussed as well.

5. Finally, it is not clear to me what exactly would be the benefit of the proposed method. Quantifying megacity emissions is of course a valid goal, but this could be done with simpler methods as well. So the "extra" of the proposed method would be the generation of spatially resolved emission maps. For this purpose, a discussion of uncertainties and "reliability" of emission values for individual pixels is required. In addition, the authors should indicate concrete applications for the derived emission maps.

Thanks to the referee for this comment and raising up the question. This study proposes a methodological simple approach to obtain the emission strengths and their spatial patterns. The simplicity is one of the benefits. The method is based on simple assumptions and real observations and doesn't address the complex inverse modelling approaches. The applications are multiple, one of the most important is to determine the location and quantification of emission sources. Since it's based on actual observations, it helps to support the development and implementation of the control and mitigation policies.

We agree that several simplifications (as commented in Q4) introduce uncertainties. However, the use and benefits of this new approach are also shown. This method does quite a good job at identifying areas which have high emissions variability, and/or were missed by previous studies. This would likely be of use in regions without detailed bottom-up inventories such as in the Global South, or in areas with a rapidly changing emissions profiles, such as those undergoing significant biomass burning.

We have added another section to investigate the uncertainties. Consideration of seasonal and spatial variations of α and τ will be future target, which might help to reduce the biases.

Further comments:

6. Selection of sample locations: application of the method for Riyadh is quite straightforward due to the good observation conditions, as well as the split of wind directions almost equally in two opposite directions. But I wonder how the method should work for Madrid, as there is basically one dominating wind direction. So the wind-assigned anomaly can definitely tell you where Madrid is located, but with this "unimodal" wind pattern, I really wonder what additional information on spatial distribution of sources should be gained. There might be other megacities where the approach might be more promising.

The uncertainty of the wind segmentation has been discussed in Section 4.2. Approximate 12.5% and 8.6% changes are observed if we use a different wind segmentation of NE/SW for Riyadh and SE/NW for Madrid, respectively. The spatial distribution of estimates in Riyadh changes slightly when wind segmentation changes. However, some positive emissions are obtained in southwest of Madrid. Using different wind segmentation leads to different spatial distributions of estimates, especially in Madrid where the topography (e.g., land cover, altitude) is more complicated than in Riyadh.

We selected Madrid to confront our proposed method with a more problematic scene (due to the existence of a dominating wind direction, as the referee correctly points out). There are already several studies on Madrid (Bauwens et al., 2022; Levelt et al., 2022), which can be used as "reference" for evaluating the success of our method.

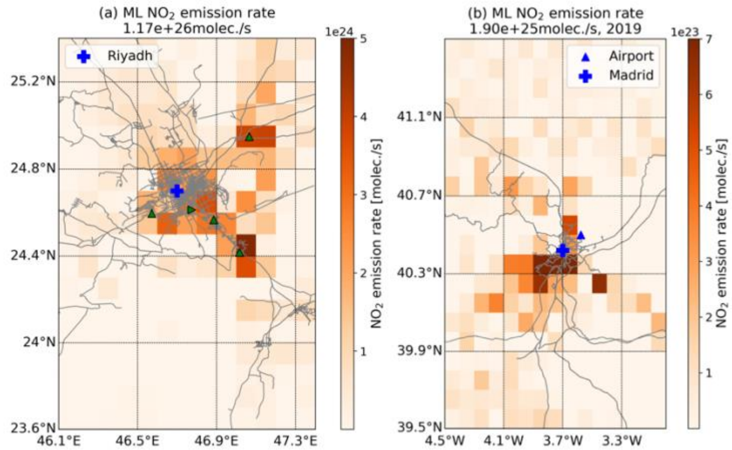


Figure 8: (a) similar to Figure 1 (d), but using SW-NE wind field segmentation for Riyadh; (b): similar to Figure 2 (d), but using SE-NW wind field segmentation for Madrid.

7. Line 70: "used to train": training is a crucial element of any ML, and I wondered here against which "truth" the ML is trained. It needs Eq. 1 to understand how the "modeled truth" is constructed that is used for training. I think that this is also an important component of this approach, that a simple downwind plume model is used to construct the NO2 distribution for the emissions from each grid pixel.

In this method, the true values (also called labels) are the TROPOMI observations. The referee is right that the “modeled truth” is used for training and obtaining the $\boldsymbol{w}$, i.e., the emissions in the end.

We cannot train the true values and the sentence is wrongly stated. It has been changed to:

“In this study, the Gradient Descent (GD) approach in ML incorporating the wind-assigned method (Tu et al., 2022a, 2022b) is used to train the “modeled truth” constructed from a simple downwind plume model for the emissions on each grid pixel using space borne NO_2 observations, to estimate the NO_2 emission strengths of two (mega)cities: Riyadh (Saudi Arabia) and Madrid (Spain)”.

8. Section 2.2: The authors describe eq. 1 as the "averaged distribution ... over a long time period" (Line 90), but later apply this to "daily plumes". Please clarify.

Eq. 1 might introduce biases for a single day as the choice of opening angle is either too small or too large, or lifetime is either too short or too long, or the distribution of emitted NO_2 is not even. However, this equation becomes reasonable for a long-term period as the biases are largely compensated.

9. Eq. 1: what should be the meaning of the division by an angle in degree? I think this cannot be correct - is there a sin or cos missing? Otherwise, please clarify the units of all components of Eq. 1.

The opening angle of the plume has a unit of rad. We have updated this in the text.

The units of all components are clarified according to referee’s comment.

10. Line 115: I understand the motivation for choosing log(wk) as proxy for wk. However, this drastically modifies the weight of the different pixels with focus on very low emission values. The satellite measurements, on the other hand, have highest signal to noise for the pixels with very high emission values.

This has to be discussed. Have you tried to run the algorithm directly with wk instead? This will of course result in some negative emission values, but the results for strong sources like powerplants might be more reliable.

Please also specify the exact procedure: are Eqs. 5-7 applied for wk or actually for log(wk)? If the latter is the case, then there should also be log(wk) written in all equations, plus an additional equation indicating how final emissions are derived (perhaps trivial, but I think very helpful for understanding what was done exactly).

Thanks for this important comment. We have tried to run the algorithm directly with wk, i.e., linear emission (see Figure R7). The spatial patterns of estimates (except the negative ones) are comparable to our previous results. The total emission rates decreased by 30% for Riyadh and 2% for Madrid, respectively. More negative emissions are modelled for Riyadh in south region than for Madrid. This means, using log(wk) does not drastically modify the weight on the pixels with low emission values.

We believe the log(wk)-based reconstruction is superior despite of assigning higher weights to small positive emissions, because it excludes unphysical negative emissions, which occur in the linear reconstruction.

log(wk) is just a proxy in the ML process, but the real wk, i.e., exp(log(wk)) is used in Eqs. 5-7.

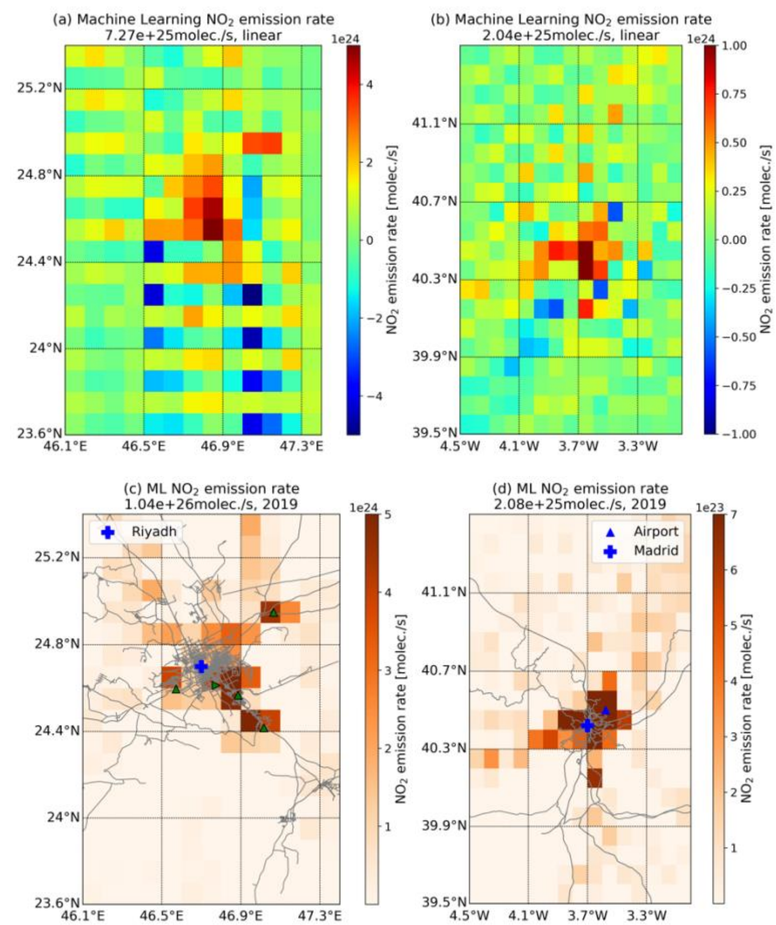


Figure R7: Estimated emission in Riyadh and Madrid without using log for wk (a-b) and using log (c-d) (results based on data in 2019).

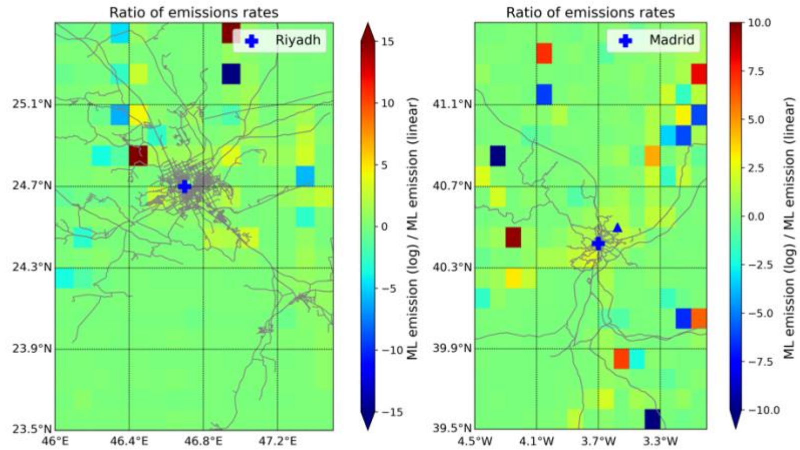


Figure R8: Ratio of estimated rates between using log and linear for Riyadh and Madrid.

11. Line 115: Where is the initial epsilon coming from?

The initial epsilon is an empirically random value, or in other words it can be any number. We use 1E26 here. This value will be later scaled by the weights w.

12. Line 132: If the outer ring should be skipped in order to skip edge effects, the initial study area must be (n+2)x(m+2), since one pixel on each side has to be skipped.

Thanks! We have corrected this mistake.

13. Line 152: In addition, for Riyadh the separation into two wind regimes suggests itself.

Thanks to the referee. The sentence has been added in the manuscript:

“The typical two wind regimes presented in Riyadh favors the applicability of the wind-anomaly method and is another reason of choosing it for the work.”

14. Line 279: I would not agree that the spatial patterns agree very well. There are some artefacts present in the presented emission maps, and the conclusions should reflect this.

This sentence has been changed to “The spatial pattern of the estimated emission strengths on the main sources near the city center agrees with the results from Beirle et al. (2019) as well”.

15. I'm no native English speaker myself. However, several sentences and formulations sound strange to me. I would recommend careful language check after dealing with the requested modifications/extensions.

Thanks for pointing this out. We have done our best again to improve the language.

Technical issues:

16. Line 49: etal. -> et al.

Thanks. Corrected.

17. Line 50: TROPOMI acronym explained twice.

Thanks. Corrected.

18. Line 80: To which area are the 910,000 measurements refer to? I think it would be more useful to give a typical number how many measurements there are per 0.1° grid pixel.

The information of the study areas has been added to the text. Additionally, the total amounts of measurements have been updated as the study period is extended. The sentence below has been added to section 2.1:

“There are nearly 1,380,000 in Riyadh (23.5ºN – 25.5ºN; 46ºE – 47.5ºE) and 930,000 measurements in Madrid (39.5ºN – 41.5ºN; 4.5ºW – 3ºW) of good quality over three years.”

The following figures represent the amounts of TROPOMI NO_2 measurements in each 0.1º grid pixel and they are added to the appendix. The following sentence has been added to section 2.1 as well:

“The amounts of TROPOMI measurements in each 0.1º grid pixel is distributed evenly with a number range of 4500-5000 in Riyadh, whereas larger difference is observed in Madrid with a number range of 2300-4000 (see Figure A- 1).”

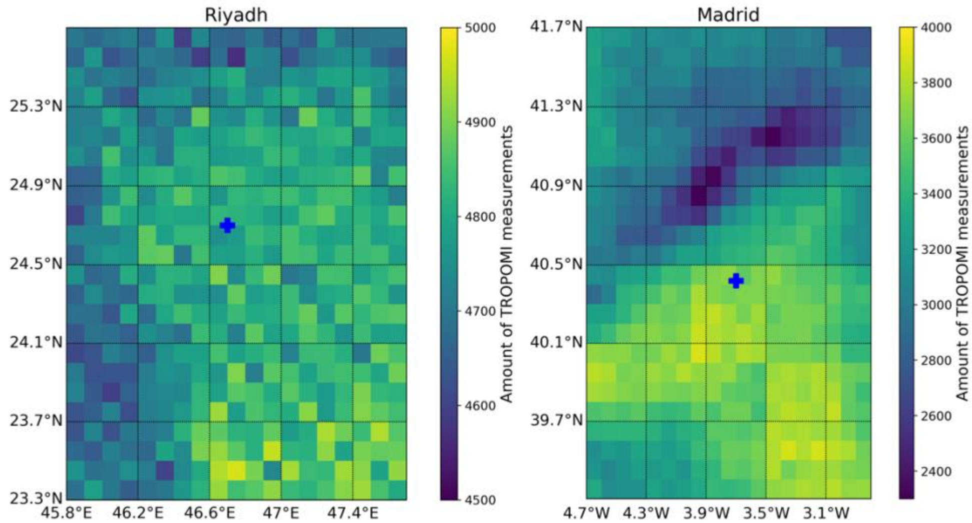


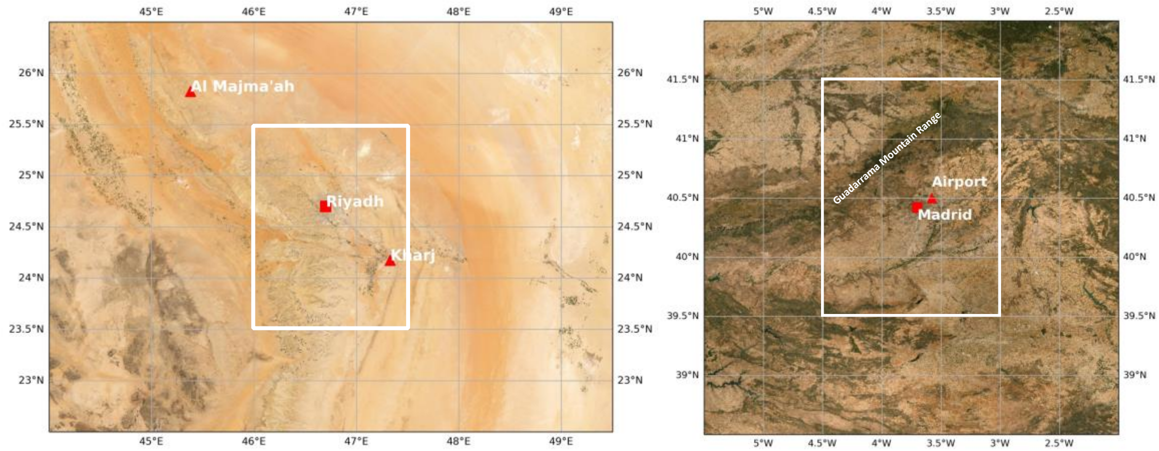
Figure A- 1: Amount of TROPOMI measurements in each 0.1° grid pixel for Riyadh and Madrid during May 2018 – August 2022.

19. all emission maps: The emissions are given in molecules per second, but implicitly refer to the chosen grid. I.e., for 0.05° grid, numbers would be only 1/4 of the presented values. Emission maps should thus be provided as densities (emissions per time per area).

Thanks for this comment. The referee is right that emission densities would be grid-independent and perhaps a superior choice, but we believe we have handled the required scaling of emissions per grid box when transitioning between the two different grid sizes correctly.

20. all maps: please choose the same lat/lon range for all plots for Riyadh and Madrid, respectively.

The terrain maps have been updated by adding latitude and longitude to the figures (as suggested by the other referee). Larger areas are preferred as it can help readers to better understand the topography of the whole area. The study areas are illustrated by using white rectangles. It is difficult to specify the boundary of the study area in the zoom version for Madrid in Figure A- 11. We would like to keep this figure for showing the land use of the Madrid as an additional information.



21. Caption Fig. 1: 0.1° is coarser than most TROPOMI pixels, thus the data is not "oversampled".

The "oversampled" seems bringing misleading information. The sentence has been changed to "Data in (a), (b), and (d) are gridded on a regular latitude-longitude grid with 0.1° spacing".

22. x labels of Fig. 1 (c), 2 (c), 3 (f): "TROPOMI tropospheric NO2" is misleading here, as the shown quantity is a difference (or anomaly).

The referee is right that the x labels in these figures represent the wind-assigned anomalies derived from TROPOMI tropospheric NO_2. This information can be obtained from the subtitle of the figures. The x and y labels represent the data sets from where the wind-assigned anomalies are derived. We have added this information to the figures' caption.

23. Figs. 3 (c) and 4 (c): There are several pixels where weekend emissions are higher than on weekdays. Thus the colorbar in (c) should be symmetric around 0, including negative values as well.

Thanks. The figures have been updated.