Author Responses to Reviewer Comments

We thank the reviewers and the editor for their useful comments and feedback. We have reproduced the reviewer comments below in black text, followed by our responses in red text. Any additions to the manuscript are in blue text and our reference to line numbers is based on the original manuscript submitted. Please note that as several of the comments from the reviewers are similar and our responses interlinked, we have produced one response document containing all the reviewer comments and our responses. We will upload this document in response to each set of comments received on-line in the ACPD discussion stage.

Reviewer 1's Comments:

While the topic is generally of high interest and would match the scope of ACP, I recommend to reject the paper due to methodological flaws and shortcomings, and as I do not see the claim of the title supported by robust results.

Reviewer 1 has provided some useful and constructive comments which will help us improve our manuscript. However, while the reviewer has suggested rejection, we have provided detailed responses below, many of which do not require major changes and are relatively straight forward to address, which we believe should make our manuscript suitable for publication in ACP.

3.1 NOx sources

Spatial correlation between major NOx emission sources and TCNO2 could be already seen in GOME data in 1996 – this is neither new nor surprising. But as surface emissions and TCNO2 are different quantities (as stated by the author), it is hard to derive any quantitative information from this. For me the conclusion of Fig. 3 would be that there is no obvious discrepancy between the spatial distribution of NOx sources in NAEI and TCNO2, but I don't see how far this comparison might help to improve the bottom-up inventories.

While this sort of analysis has been done before, the use of TROPOMI compared with GOME is much more useful given the much higher spatial resolution of TROPOMI. Given that TROPOMI has a horizontal footprint of 3.5 km x 5.5 km, compared with 320 km x 40 km for GOME (Burrows et al., 1999), it has approximately 665 pixels to every one of GOME's pixels. Therefore, while GOME is able to provide information on a regional basis (e.g. 1 or 2 pixels for South East England), TROPOMI is able to resolve English city and town NO_2 sources making it a much more appropriate instrument to perform this satellite-emission correlation analysis. Also, as the reviewer has not provided a reference for the 1996 GOME NO_2 – Emissions NO_x correlation analysis, it is difficult to comment directly on those results in relation to our study here. While the reviewer is correct that the result of Figure 3 is that there are no obvious discrepancies between the two quantities that in itself is an important result as it provides confidence in the spatial distribution of the NAEI emission inventory. For instance, if there had been some obvious missing sources, this analysis would have likely determined it. However, in line with the Editor's comments, we have changed the manuscript title to "Exploiting satellite measurements to explore uncertainties in UK bottom-up NO_x emission estimates".

3.2 Trends

The authors compare trends in NOx emissions and TCNO2 for 2005-2015. However, the row anomaly in OMI started already earlier than 2015 and potentially degrades the OMI timeseries in a nonlinear way. Thus, I do not consider the presented trend estimates from OMI TCNO2 to be robust.
The discussed trends in AURN for urban/background/rural sites might be related to nonlinear chemistry. But they might also be just related to different emission trends for different sectors or regions, which is not resolved in the presented NAEI data (though the information would be available, I assume).

Yes, the row anomaly did start before 2015, but the OMI $TCNO_2$ is still reliable across the UK up until approximately 2015, as shown in the study by Pope et al. (2018). And while the row anomaly after 2015 starts to substantially influence background $TCNO_2$, the large source regions still show sensible signals and can be used for trend analysis for later years as shown by Vohra et al. (2021). Therefore, we respectfully disagree with the reviewer and suggest that the $TCNO_2$ trends presented in this study are suitable for comparison to that of the NAEI NO_x emissions. However, we have updated the sentence "We did not consider OMI $TCNO_2$ after 2015 as the row anomaly substantially degraded the quality of the data over the UK from this point." to "The OMI row anomaly first occurred in 2008 (Torres et al., 2018) and over time has progressively had a detrimental impact on retrieved $TCNO_2$. The study by Pope et al., (2018) successfully used the OMI record to look at long-term trends in UK $TCNO_2$. However, after 2015, while still retrieving robust signals over source regions, the row anomaly appears to be substantially artificially enhancing background $TCNO_2$. Therefore, as we consider regional trends in $TCNO_2$ in Section 3.2, we did not use OMI $TCNO_2$ after 2015.".

As for the AURN urban background/suburban/rural sites, they measure NO_2 from nearby sources which will be from a range of sectors (most notably from transport and energy production). The rural sites are also substantially fewer in number than the urban background/suburban sites. Therefore, trends in surface NO_2 should be a reasonable representation in changes in anthropogenic NO_x emissions. Also, as annual means are used, this will average out much of the meteorology which influence the atmospheric chemistry of NO_2.

3.3 Top-down emissions

The authors state that they follow a simple approach in order to derive NOx emissions from TCNO2, and apply a formalism modified from similar methods in literature. While I see the potential of the general approach for detecting and quantifying NOx emissions from megacities on global scale, I consider it quite challenging to quantify NOx emissions of smaller cities with an accuracy good enough to be of use for improving bottom-up inventories of such high quality as NAEI.
For the detailed implementation, I see the following shortcomings:

(a) I don't understand the justification for the method used for determining the background. Why should the intersect of the wind-flow LD with the all-flow LD be a "reasonable estimate" for this? As shown in Fig. 1b, wind-flow and all-flow LDs are complex curves with several minima and maxima. Searching for intersects just yields a number of quite arbitrarily distributed points. Why is B not just included as additional parameter in the fit?

Firstly, we do not use the method of where the wind-flow LD with the all-flow LD profiles intersect to determine the background level. We do state that this would be a "reasonable estimate", but as shown in Figure 2, broad large-scale background enhancements can mean the two profiles never intersect in the domain we are using. In a location with an isolated source, this approach would be more suitable because the wind-flow profile downwind of the source should be larger than that of all-flow profile and when they intersect it is indicative of where the wind-flow profile has returned to average (i.e. background levels). However, as this is not clear in the manuscript, we have updated the text on Line 186 to "A reasonable estimate of when the wind-flow *NO_2 LD* reaches *B*, for more isolated NO_2 sources, is when it intersects with the all-flow *NO_2 LD* profile (i.e. returns to normal levels).".

As to fitting the background parameter, the approach in Berlie et al. (2021) would not be suitable given the larger frequency of NO_2 sources in the UK, so the approach of Verstraeten et al. (2018) which accounts for upwind background influences would be more applicable. However, while we derive the background NO_2 value offline, before using Eqn 1, we respectfully disagree with the reviewers and believe that the T-Test approach is still an appropriate methodology to derive B. In the comment "*As shown in Fig. 1 b, wind-flow and all-flow LDs are complex curves with several minima and maxima. Searching for intersects just yields a number of quite arbitrarily distributed points*", the reviewer is correct that some of the downwind profiles can be complex. In the case of London in Figure 1, the downwind sources are from continental Europe. However, we are not using the approach where the wind-flow and all-flow regimes intersect which is clearly state in Section 2.3. The T-Test approach successful detects where there is either a turning point in the profile (i.e. a background level before another source in the downwind profile is found or where the gradient in the NO_2 LD has plateaued. Once the e-folding distance has been determined, based on the profile between the source and B, the corresponding NO_2 lifetime determined in the study e.g. for London (3.8-8.3 hours) and Edinburgh (5.2 hours) are comparable to that of Verstraeten et al. (2018) for London (4.6+2.5 hours) and Edinburgh (4.2+1.8 hours). Therefore, providing confidence in the methodology used here to derived, B, the e-folding distance and the lifetime.

(b) Eq. 1 is a rough simplification; it does not account for the spatial extent of the city, and it completely ignores additional emissions. In particular for cities like Manchester, with several smaller cities in the East, this will affect the lifetime estimate for the westerly wind case, and thus also the estimated emissions are affected.

Eq. 1 accounts for the width of the source in the calculation of the NO_2 line density. The NO_2 line density is calculated as (new equation 2):

$$NO_2LD_{i=1,N} = \frac{\sum_{j=1}^{n} TCNO_{2i,j}}{n} \times w$$

where *NO_2 LD* is the NO_2 line density, *i* the grid box index downwind of the source starting at *i=1* going to *i=N* at background point *B*, *$TCNO_2$* is the tropospheric column NO_2 grid box value at point *i* and *j* is the grid box index for the number of grid boxes *n*, perpendicular to the downwind profile, which fit across the width of the source at grid box *i* downwind and *w* is the width (i.e. source width perpendicular to the downwind profile) of the NO_2 source and *α* is the grid box width of the

downwind grid box at point *i*. To make this clearer, this equation and corresponding text have been added to the manuscript on line 174 and replaces lines 174 to 181. The text now reads:

"The NO_2 *LD* is the product of the source width, which is perpendicular to the wind flow, and the source-width-average $TCNO_2$ profile downwind from the source on a grid box by grid box basis as shown in Equation 2.

$$NO_2LD_{i=1,N} = \frac{\sum_{j=1}^{n} TCNO_{2i,j}}{n} \times w \quad (2)$$

where NO_2 *LD* (moles/m) is the NO_2 line density, *i* the grid box index downwind of the source starting at *i=1* going to *i=N* at background point *B*, $TCNO_2$ is the tropospheric column NO_2 grid box value (moles/m^2) at point *i* and *j* is the grid box index for the number of grid boxes *n*, perpendicular to the downwind profile, which fit across the width of the source at grid box *i* downwind and *w* is the source width (m) (i.e. source width perpendicular to the downwind profile) of the NO_2. Though the source width is a subjective choice between the source edge locations, the same source width value is used when deriving the TROPOMI NO_x emissions and summing up the NAEI NO_x emissions over the source region. As the source emissions will be a function of the source width (i.e. larger at source centre and lower at source edge), the mean $TCNO_2$ downwind profile is representative of the source-average NO_2 emission.".

The method we employ here is dependent on a clearly defined downwind plume being detected which is not substantially influenced by downwind sources. When there are broad background enhancements from upwind sources (e.g. Birmingham easterly flow example in Figure 2), the T-Test accounts for this where the $TCNO_2$ gradient, with respect to distance, in the source plume, superimposed on the background enhancement, tends to 0.0. Therefore, some sources only have one or two wind directions due to the complicating factors listed above. In the case of Manchester emissions under westerly flow, the reviewer is correct the downwind sources from Leeds are potentially influencing the emission rate derived for Manchester and this has been removed from our analysis. The corresponding text and table have been updated.

(c) Seasonal effects (on emissions and lifetimes) are not discussed at all.

The reviewer makes a good point here. Unfortunately, the TROPOMI record is only a few years old having been launched in October 2017. Furthermore, the UK is frequently influenced by cloudy weather which restricts the use of much of the data. Therefore, the time record to derive seasonal emissions is limited so we are restricted to annual average analysis (i.e. composite $TCNO_2$ data from wind flows in all seasons). There are also several comments e.g. penultimate sentence of Reviewer 1's comment 3.4 and M. Pommier's comment 5 asking why we do not use the 2019 data directly to compare with the 2019 NAEI NO_x emissions. While the one year of 2019 does provide a clear annual mean $TCNO_2$ map over the UK, it further restricts the frequency of observations to derive emissions. Therefore, in this study we use two full years of TROPOMI data between Feb 2018 (when data is first available) and Jan 2020. Ultimately, it might have been better to use three years of data centred on 2019, but unfortunately, we cannot exploit data beyond Feb 2020 because of the COVID-19 pandemic which resulted in a large step change in NO_x emissions and is not representative of normal conditions. Also, seasonality has an impact on the frequency of observations and their contribution

to the annually derived top-down NO_x average. To address these points we have added the following paragraph to the end of Section 2.3:
"In this study, the top-down NO_x emissions are derived by sampling $TCNO_2$ data under different wind directions in all seasons. Several studies, such as Beirle et al. (2011), have gone a step further and used $TCNO_2$ data to derive seasonal emissions. Unfortunately, here we are restricted to looking at annually derived emissions due to 1) the TROPOMI $TCNO_2$ record only started in February 2018, 2) the COVID-19 pandemic resulted in a dramatic reduction in UK (and global) NO_x emissions (Potts et al., 2021) meaning $TCNO_2$ data beyond February 2020 could not be used to derive top-down emissions under normal conditions and 3) the UK is subject to frequently cloudy conditions yielding a reduction in the number of observations from TROPOMI. The latter point predominantly influences TROPOMI retrievals in the winter-time. Therefore, even though we sample $TCNO_2$ data in all seasons, there is likely to be a tendency towards summer-time $TCNO_2$ values, when $TCNO_2$ values tend to be lower (e.g. Pope et al., 2015), potentially leading to a low bias in the derived top-down NO_x emissions.".

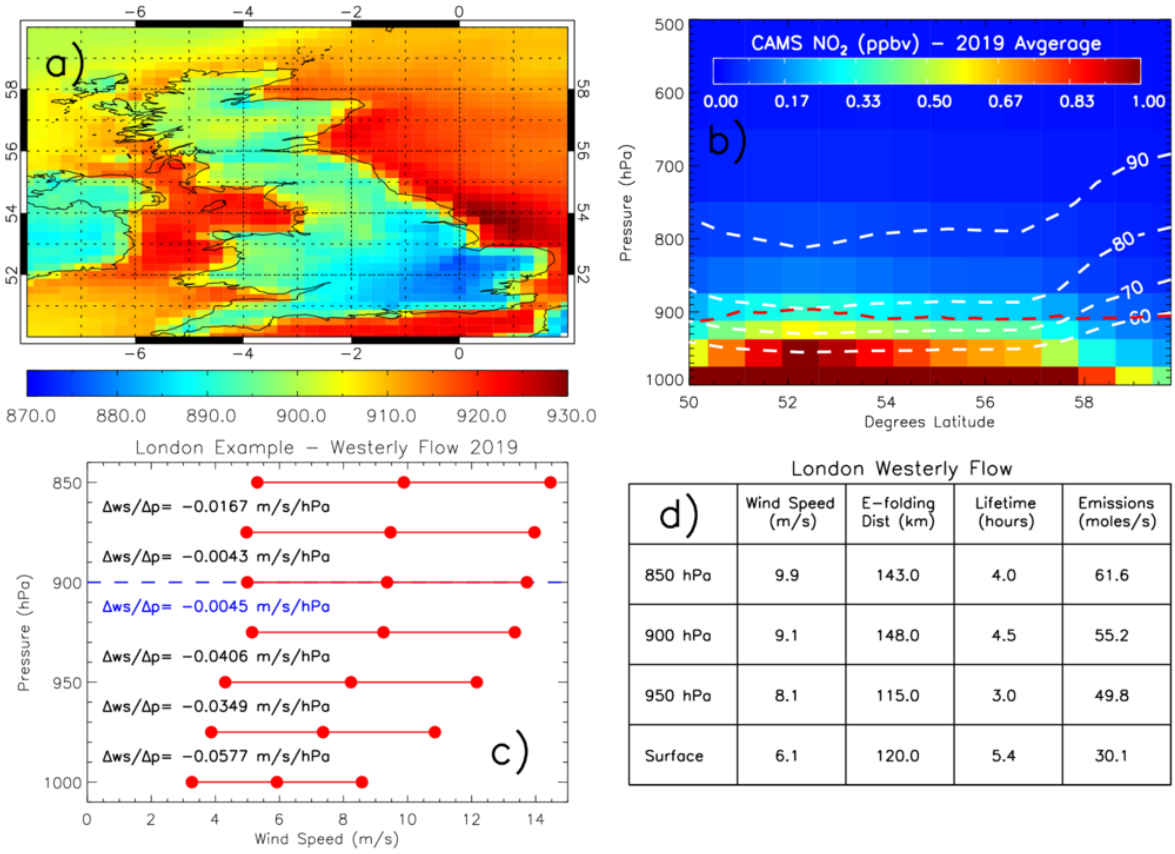
(d) Some a-priori choices (like for the "source width") are not provided in the text, and uncertainties for these choices (like selection of pressure level for wind data) are not discussed.

The source width is selected based on outer limits of the city edges, perpendicular to the flow direction. While these choices are subjective, the same source width is used when deriving the TROPOMI NO_x emissions and summing up the NAEI NO_x emissions over the source region. This has been included in our text addition to Reviewer 1's comment 3.3b.

As for the pressure levels on which to select the wind speed data, the bulk of the emitted NO_2 is within the boundary layer. Therefore, using an average wind speed through the boundary layer is a reasonable estimate, as it will be these winds which are transporting the bulk of the emitted NO_2. Here, we have taken the boundary layer height to be approximately 850 hPa. Both Beirle et al. (2011) and Verstraeten et al. (2018) used an upper altitude of 500 m, which is approximately 950 hPa over the UK. Beirle et al. (2011) did investigate the uncertainty of this choice (~30%), but neither study provided a detailed reasoning for selecting 500 m as the upper limit. Therefore, we have added a new figure to Section 2.3 to investigate the most appropriate level to select winds from and average to the surface.

From the new Figure 1a, the boundary layer pressure ranges between approximately 880 and 910 hPa for the 2019 average. This is approximately 1000 m as boundary layer height. Therefore, this suggests that using 850 hPa as the upper limit to average over is too large an altitude range. From Figure 1b, the bulk of the NO_2 loading in the UK, in the zonal 2019 average, is in the lower troposphere with approximately 60-70% of the surface-500 hPa NO_2 loading between the surface and 900 hPa. When the zonal mean boundary layer pressure is overplotted, this overlaps the 900 hPa further supporting the conclusion that the surface to 900 hPa layer is the most representative wind speed average to use as it is the winds across this altitude range which will be controlling the bulk of the NO_2 spatial distribution from sources. Figure 1c shows a wind speed profile for London under westerly flow. At each pressure level, the u- and v-wind components (averaged between the surface and pressure level) are used to determine the westerly flow direction. Therefore, each derived wind speed average for each level can have different sample sizes. From the profile, there is

a clear increase in wind speed with altitude until about 900 hPa. Here, the wind speed gradient between pressure levels drop from -0.0406 m/s/hPa between 950 hPa and 925 hPa to -0.0045 m/s/hPa between 925 hPa and 900 hPa (i.e. blue colouring) signifying less turbulence (i.e. near the top of the boundary layer) and the flow becomes more laminar with altitude. The table in Figure 1 shows the different parameters for the NO_x emission calculation for London under westerly flow. When deriving the average wind speed between the surface and 850 hPa, the NO_x emission is 61.6 moles/s. However, when using the surface winds, this drops to 30.1 moles/s. At 900 hPa and 950 hPa, the emissions are 55.2 and 49.8 moles/s. Therefore, between 850 hPa and the surface, there is a substantial change in the NO_x emission rate. However, the step change between the other levels is less dramatic (i.e. above 950 hPa the rate is between approximately 50.0 and 62.0 moles per second). Therefore, we suggest that the surface to 900 hPa, not surface to 850 hPa as in the original manuscript, is the most appropriate pressure range to average over.



London Westerly Flow

d)	Wind Speed (m/s)	E–folding Dist (km)	Lifetime (hours)	Emissions (moles/s)
850 hPa	9.9	143.0	4.0	61.6
900 hPa	9.1	148.0	4.5	55.2
950 hPa	8.1	115.0	3.0	49.8
Surface	6.1	120.0	5.4	30.1

Figure 1: a) ERA-5 UK boundary layer pressure (hPa) sampled at 13.00 LT (to coincide with the TROPOMI overpass time) and averaged for 2019. b) CAMS reanalysis zonal (8.0°W-2.0°E) average latitude-pressure NO_2 (ppbv) cross-section over the UK between the surface and 500 hPa. White dashed lines represent the percentage of the surface-500 hPa NO_2 loading between the surface and the respective pressure levels. The red dashed line represents the zonal average boundary layer pressure (hPa). c) Average (surface to pressure level) wind speed (m/s), ± the standard deviation, profile over London under westerly flow (determined from the ERA-5 u-wind and v-wind components at each pressure level). Δws/Δp is the wind speed gradient between pressure levels. The blue text indicates the first small step change in the gradient indicative of reduced flow turbulence and a suitable surface-altitude range to average the winds speeds over. d) The table

shows the impact to the NO_x emission parameters when using different altitudes over which to average the wind speeds.

At line 172 we have split the text into two paragraphs. The original text "and averaged across boundary layer pressure levels (i.e. 1000 hPa and 850 hPa)." on line 170-171 has been removed and replaced with "Studies such as Beirle et al. (2011) and Verstraeten et al. (2018) averaged the wind speeds over the surface to 500 m layer. Beirle et al. (2011) suggested that the average winds across this altitude range yielded uncertainties over approximately 30%, but neither study provided definitive reasoning why 500 m was selected. In the UK, 500 m is approximately 950 hPa which sits comfortably within the boundary layer (approximately 1000 m or 880.0 to 910 hPa in **Figure 1a** based on ERA-5 data sampled at 13.00 LT and averaged for 2019). In this study, we argue that wind speeds throughout the boundary layer are likely to be important in controlling the spatial distribution of NO_2 downwind of sources. **Figure 1b** shows the zonally averaged latitude-pressure NO_2 profile from the Copernicus Atmosphere Monitoring Service (CAMS), sampled at 13.00 LT and averaged for 2019, over the UK. The bulk of the NO_2 loading is near the surface with NO_2 concentrations of 0.5 ppbv to >1.0 ppbv between the surface and 900 hPa. As shown by the white dashed lines, 60-70% of the surface to 500 hPa NO_2 loading exists between the surface and 900 hPa. The zonally averaged boundary layer pressure (red dashed line) also straddles the 900 hPa level. In **Figure 1c**, the wind speed profile for London sampled under westerly flow increases with altitude until between 925 hPa and 900 hPa. For each pressure level, London westerly days are defined based average u- and v-components between the surface and the respective pressure level. As shown by the blue text, the wind speed gradient with respect to pressure substantially decreases (i.e. from -0.0406 m/s/hPa between 950 hPa and 925 hPa to -0.0045 m/s/hPa between 925 hPa and 900 hPa) at 900 hPa. Therefore, this profile gradient and the information in **Figures 1a** & **b** suggest that 900 hPa is a suitable level to derive the boundary layer average wind speed and flow direction. The table (**panel d**) in **Figure 1** shows the sensitivity of the NO_x emission parameters to the pressure layer used. The surface-850 hPa average and surface only winds show substantially different NO_x emission rates of 61.6 moles/s and 30.1 moles/s, respectively. However, the intermediate levels (900 hPa and 950 hPa) show less dramatic step changes with emission rates of 55.2 moles/s and 49.8 moles/s. Therefore, the surface-900 hPa layer is used to help derive NO_x emission rates in this study.".

We have added the following sentence to the data section for the CAMS data "CAMS NO_2 data was retrieved form https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-reanalysis-eac4?tab=form.". We have also updated the sentence for the meteorological data to "Meteorological wind, temperature and boundary layer height data came from ECMWF (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview)." to include temperature and boundary layer height information as well. We have also added a reference for CAMS (i.e. CAMS, 2021) in the additional reference list at the end of the document.

3.4 Comparison to GEOS-Chem

The authors compare TROPOMI TCNO2 for 2019 with a model run based on 2016 emissions. The emissions have been adjusted to 2019 using the linear trend derived from the temporal evolution of NAEI total emissions 2009-2018. However, this approach does not account for sector- or region

specific changes. For instance, it could be that the decrease is mostly due to emission reduction in traffic sector (thus mostly affecting London emissions), while rural emissions stay more or less constant. Thus, the spatial patterns shown in Fig. 6 might actually be also caused by a change of emissions in 2019 compared to 2016 which cannot be just described by an overall linear decrease. Overall, the authors derive emission estimates and trends from satellite observations of NO2 based on rather simple assumptions, and compare those to bottom-up inventories. Despite the methodological shortcomings listed above, the authors find rather good agreement to NAEI bottom up emissions. Basic conclusions, however, are in essence that no disagreement was found.
But even with modified/improved methodology, I doubt that the approach presented in this study, applied to 1 year of TROPOMI data, would allow to actually improve a bottom-up inventory for the UK. For this, I would expect that more refined methods and consideration of longer time periods would be necessary.

In line with this comment from Reviewer 1, Reviewer 2's comment 1 and the Editor's suggestions, we have removed the GEOS-Chem analysis from the paper. Though it would be useful to rerun the model using the 2019 emissions, unfortunately it is now beyond the scope of this work to do this.

Additional comments:
- the maps of UK have a strange aspect ratio; it seems to me that the projection was trying to correct for the latitude grid being shorter than the longitude grid at 53°N, but applied the correction in the wrong direction and stretched the latitude aspect instead of squeezing it.

To be able to clearly show the city-scale sources being detect by TROPOMI, we had to zoom in on a regional basis yielding this projection. While not ideal, we argue that the lons and lats are labelled so the reader can easily identify the location and that the current view clearly shows where the sources are with the relevant city labels overplotted.

- figures are hard to read, in particular Fig, 3 a-d.

We have made the text larger in this figure.

Reviewer 2's Comments:
This paper presents satellite measurements of NO2 from OMI and TROPOMI for the UK, and estimates NOx lifetimes, emissions, and trends. The goal is to reduce uncertainties in bottom-up emission inventories. The paper itself is well written. However, some figures need updates (legend, too small caption, …), some concerns need to be verified:

1. When comparing the results with the GEOS-Chem model, emissions are derived from 2016 emission sources. This does not take into account sector- or region specific changes. If possible, this should be reconsidered.

We have now removed the GEOS-Chem analysis from the manuscript. Please see our response to Reviewer 1's comment 3.4 for more information.

In order to highlight the differences between the TROPOMI output and the chemical transport model, it would be good to consider a higher resolution nesting with the GEOS-Chem model.

While this is an interesting pointed raised on using a higher resolution nesting for GEOS-Chem by the reviewer, this would be a very large task and as the model is an aside to the main objectives of the manuscript, we argue that is beyond the scope of this study to implement this, while likely returning very similar results.

2. If you use ECMWF wind data from ERA5 for the SMBA, why not considering this model as input for your GEOS-Chem model?

As in response to Reviewer 2's comment 1, the model is a secondary addition to the paper and any large-scale changes to model setup will be extremely time consuming. GEOS-Chem by default uses NASA Global Modelling and Assimilation Office (GMAO) Goddard Earth Observing System – Forward Processing (GEOS-FP) assimilated meteorology, which like ERA5, is a reanalysis product, meaning that the meteorological fields will be highly consistent and will not yield any substantial benefits to that of the existing GEOS-Chem runs.

3. The use of the quality flag is not clear to me, can you please explain which threshold value you are using here?

This has been addressed in our response to M. Pommier's comment number 5. Here, we have now reprocessed the TROPOMI data using a QA flag >0.75 instead of a QA flag > 0.5.

M. Pommier Comments:

However, I have major concerns in the results published in this work.

1) The use of the NAEI16 should be discussed and explained a bit more by the authors. It is a 3-yr old dataset and the authors have decided to interpolate this emission inventory to represent 2019.
The NAEI19 is available as explained in the text. I also understand it was not feasible to use this new inventory at the beginning of the study. But there is no explanation on the interpolation used.

Since the NAEI19 is available (and other more recent years too), a minimum should be to check and present an evaluation of their estimated 2019 emission in comparison to the official NAEI19 and see if their assumption is correct or not. I guess it will not change the main conclusion of the paper, but I think the assumptions used by the authors are critical and need more explanations.

We have replaced the 2016 NAEI emissions with that of the 2019 version and updated the manuscript accordingly.

2) The authors should also justify their criteria used for the TROPOMI data. They used a quality flag of 0.5 with additional criteria. However, the ESA does not recommend using the data with a quality flag lower than 0.75. This higher value in the quality flag avoids the cloudy scenes and the problematic retrievals.
Loads of studies use this flag at 0.75. To cite a few, there are:
- Bierle et al. 2019 Science: https://www.science.org/doi/10.1126/sciadv.aax9800
- Zhao et al. AMT 2020: https://doi.org/10.5194/amt-13-2131-2020

- Ialongo et al., Atmos. Env. 2021: https://doi.org/10.1016/j.aeaoa.2021.100114
- Lange et al. ACPD 2021 : https://doi.org/10.5194/acp-2021-273
Moreover, there is the change in the cloud filter in 2019 which impacts the data with a quality flag of 0.5. Thus, I think the TROPOMI values presented in this study are overestimated.

We have now reprocessed the TROPOMI $TCNO_2$ data with a qa flag >0.75.

3) The method used by the authors to estimate the NOx emissions is sensitive to several parameters, such as the BL height (they have chosen the levels between 1000 and 850 hPa) and the coordinates of their point source (city). A sensitivity analysis could be done, or at least a discussion should be added.

This is addressed extensively in our response to Reviewer 1's comment 3.3d.

4) I do not see a clear conclusion about the results given for the different wind directions. If I take the example of London, the emission rate varies from 36.60 to 58.70 mole/s.
The result based on the S direction for London is not presented (Table 1). Is it because there is not enough data to have a representative estimate? Based on the sentence in line 400, I guess the S wind direction for London does not frequently occur. I have the same comment for the other cities, why did the authors provide an estimate only for one or two wind direction(s)? Their method seems very sensitive to the wind direction used.

For several of the cities we investigated we were unable to derive a top-down NO_x emissions estimate and for others it was only certain wind directions where we could derive an emissions flux. Multiple reasons include the lack of data (mainly due to cloud cover), complex upwind sources and the size of the targeted source being too small for TROPOMI to detect substantial enhancements. Therefore, a clear point we did not make in the original manuscript is that it is very challenging in the UK to derive top-down emissions and that we are one of the first studies to attempt this for sources beyond London and Edinburgh (i.e. Verstraeten et al. (2018) did this using OMI). We have make changes to the manuscript to highlight this point, which are presented in our response to M. Pommier's minor comment on line 392 (below).

5) The conclusions are based on a comparison between one emission year (2016 scaled to represent 2019) with ~2 years of TROPOMI data (Feb 2018-Jan 2020). Why the authors did not use only the year 2019 for the observations? This difference should be better stated through the manuscript, i.e., in the abstract, in the figures (in the axis and the caption of Fig.5; in a legend for the maps and on the axis in Fig. 3). This gives a confusing message.

In our response, and additions to the manuscript, to M. Pommier's comment for line 392 (see below), we make it clear that cloud cover is a substantial problem when retrieving information on trace gases over the UK. Therefore, we use two years of TROPOMI data as it increases the sample size to detect robust $TCNO_2$ signals over the UK. We have added the sentence on Line 139:

"Given the issues with large cloud cover in the UK, we use two years of TROPOMI $TCNO_2$ data to help increase the spatiotemporal sample size when deriving top-down emissions to evaluate the 2019 NAEI NO_x emissions.". We have also made it clear throughout the manuscript that we are using two years of TROPOMI data (i.e. in the relevant figure captions).

Minor comments:
- All the figures are hard to read and some of them lack explanations. For example in Fig .4 , a better legend explaining the curves should be given.

We have increased the size of the text in the figures and updated the figure captions accordingly.

- In the introduction, lines 68-69, the correct sentence should be something like:
"Until recently, spatial verification of NAEI... (Tsagatakis et al., 2021)."
https://uk-air.defra.gov.uk/assets/documents/reports/cat09/2107291052_UK_Spatial_Emissions_Methodology_for_NAEI_2019_v1.pdf That's true, this report is not an evaluation of the emission maps with satellite observations, but these satellite observations are now included in the analysis.

We thank the reviewer to bringing this to our attention. This report had not been published online when we originally submitted our manuscript to ACP. We have updated lines 68-69 to:

"Spatial verification of the NAEI AQ emissions, until recently (Tsagatakis et al., 2021), has been restricted to comparisons with surface sites, which have limited and disproportional spatial coverage.".

However, having had a brief look at the report, there are a few issues which spring to mind. In the NO_2 and SO_2 sections, the authors of the report often use the term "concentrations", which is not true as these are column quantities. Also, the spatial distribution of the TROPOMI SO_2 product was unexpected. UV-Vis instruments (e.g. like OMI) are known to have difficulty detecting smaller SO_2 sources (e.g. over the UK and Europe). For instance, one would expect big point sources like Drax Power Station to show up, which is not evident here. Therefore, we suspect that the SO_2 spatial distribution in the report (similar to OMI SO_2 maps that R. Pope has looked at in the past) are more likely to be noise rather than robust SO_2 signals.

- The authors should provide the reference of the ECMWF wind data they have used.

We have included a reference for the ERA-5 data in the additional reference list at the end of this document.

- I do not understand the sentence in line 210 "…and τ of 3.8 hours (7.0 and 2.6 hours)". Does it mean the value ranges from 2.6 to 7 with a mean equal to 3.8 hours?

Yes, this is the case. In the rest of the sentence it says the wind speed is 9.9 ± 4.6 m/s (the standard deviation). This results in a lifetime of 3.8 hours at 9.9 m/s but with the wind speed range of 5.3 m/s to 14.5 m/s, the lifetime is 7.0 to 2.6 hours (i.e. if the wind speed is slower, the lifetime is longer). To make this clearer we have updated the sentence to:

"In the case of London, this yielded an e-folding distance of 150.0 km and τ of 3.8 hours (7.0 and 2.6 hours) based on the average ws = 9.9 m/s with an uncertainty range (± 4.6 m/s; i.e. ± 1-sigma standard deviation) of 5.3 m/s to 14.5 m/s (i.e. a slower/faster wind speed yields a longer/shorter lifetime).".

- Line 392: "Our methodology was applied to 12 city sources where sources had suitable downwind TCNO2 enhancements…". What does a "suitable downwind enhancement" mean? What was the criteria to choose these 12 cities?

For several of the cities we investigated we were unable to derive a top-down NO_x emissions estimate and for others it was only certain wind directions where we could derive an emissions flux. Multiple reasons include the lack of data (mainly due to cloud cover), complex upwind sources and

the size of the targeted source being too small for TROPOMI to detect substantial enhancements. Therefore, a clear point we did not make in the original manuscript is that it is very challenging in the UK to derive top-down emissions and that we are one of the first studies to attempt this for sources beyond London and Edinburgh (i.e. Verstraeten et al. (2018) did this using OMI). Therefore, we have modified the text on lines 94-105 to:

"In this study, we use satellite $TCNO_2$ records to evaluate the spatial distribution and temporal evolution of the NAEI. In the past, and still presently, this is a challenge given the climatological meteorological conditions (i.e. frequent frontal systems with widespread precipitation and cloud cover; Pena-Angulo et al., (2020)) experienced in the UK. Frequent cloud cover means that satellite instruments are severely restricted in their ability to retrieve information on trace gases and aerosols through the atmosphere (i.e. retrievals only between the top of atmosphere and cloud top). Therefore, the lack of robust observations makes it more difficult to clearly resolve large emission sources from space. Also, previous sensors (e.g. the Ozone Monitoring Instrument, OMI) have had relatively coarse horizontal spatial resolutions (in the order of 10-100 km) which are larger than most UK emissions sources. However, this work represents the first attempt to derive UK city-scale NO_x emissions from the new state-of-the-art TROPOspheric Monitoring Instrument (TROPOMI), which has unparalleled spatial resolution in comparison to previous sensors (e.g. OMI). We apply a similar approach to Verstraeten et al. (2018), but determine the background NO_2 value and e-folding distance in different ways, to derive top-down NO_x emission estimates of UK cities and thereby directly evaluate the NAEI estimates. Therefore, we can derive NO_x emissions from previously undetectable sources (e.g. Manchester and Birmingham). From here on, we refer to this methodology as the simple mass balance approach (SMBA). The satellite observations used, NAEI and SMBA are described in Section 2, the results presented in Section 3 and our conclusions discussed in Section 4.".

As to the direct point about what classifies as "suitable downwind enhancement", this is somewhat subjective. Typically, like in Figures 1a and 2a, there has to be a clear enhancement in NO_2 (i.e. positive anomalies downwind of the source) and the downwind profile has to show a clear decrease with distance so that an e-folding distance and lifetime can be determined. To make this clearer, we have updated line 392 from "Our methodology was applied to 12 city sources where sources had suitable downwind $TCNO_2$ enhancements to derive *NO_2 LDs* and top-down emissions (**Figure 5**)." to "Our methodology was applied to 10 city sources where sources had suitable downwind $TCNO_2$ enhancements to derive *NO_2 LDs* and top-down emissions (**Figure 5**). A suitable downwind $TCNO_2$ enhancement was subjectively identified when a clear $TCNO_2$ enhancement (i.e. positive anomalies) under a specific wind flow/direction occurred and a realistic lifetime (i.e. in the range of the literature – e.g. Verstraeten et al. (2018)) could be derived from the downwind $TCNO_2$ profile of the target source.".

- The last sentence of the conclusion is too general. The readers will be interested to know what is needed to improve the NOx estimates with the satellite observations, or how can we improve the NAEI with this work?

While this is a valid point, it is beyond the scope of this study. We have changed the title of the manuscript (see our response to Editor's final comment and Reviewer 1's comment 3.1) to make it clear that in the first instance the study primarily focusses on identifying what the uncertainties are. To determine what is causing discrepancies between the NAEI and top-down NOx emissions would be another paper in itself. Therefore, we feel that, while quite general, the statement is sufficient in this case.

- Figs 1 and 2. It will be interested to know the number of days and the number of pixels used in the graph. It will give an idea of the representativeness of the results.
- Fig. 5: Does the lifetime represent the mean calculated lifetime with the different estimates? If yes, it should be explained in the caption. In Table 1, the NO2 lifetime for London can be: 3.8, 8.3 and 5.3 hr. In Fig 5, it seems to be larger than 6 hours.

Yes, where there are more than one wind direction for a city which yields a top down estimate of NO_x emissions, these are averaged together in Figure 5. So for London, the calculated lifetimes are 3.8, 8.3 and 5.3 hours, which yields an average lifetime of 5.8 hours. Therefore, this has been coloured by the colour scale to near 6 hours (i.e. the dark red colour), but is not larger than 6 hours. However, the reviewer makes a good point about making this clear in the Figure 5 caption. Therefore, we have added on Line 684 "Where there is more than one top-down estimate for a city from multiple wind directions, the corresponding emission rates and lifetimes have been averaged together.".

As for figures 1 and 2, we have added text (i.e. N =) and updated the caption to state how many days of data have been used to derive the anomaly plots and then the emissions.

- Lines 423-425: "As the input emissions for GEOS-Chem come from the NAEI (2016 NAEI emissions scaled to 2019), any inconsistencies between simulated and observed NO2 potentially indicates discrepancies in the underlying emission." That's not fully true. Inconsistencies can also be due to the GEOS-Chem model (deposition, chemistry, dispersion) and/or uncertainties in the TROPOMI observations. The inconsistencies may also come from the way the authors gridded the NAEI or how they vertically distributed the emissions in the model.

While the reviewer is correct that other model processes could be influencing the model-satellite biases reported in this study, the GEOS-Chem model is a frequently used and well evaluated model. Therefore, the NAEI emissions are most likely the largest factor in the comparison discrepancies. However, as in response Reviewer 1's comment 3.4, we have removed the GEOS-Chem analysis from the study.

- Note the weblink: http://www.temis.nl/airpollution/no2.html does not work (checked on 16 September).

This link has changed since the manuscript was originally written and we have now updated it to https://www.temis.nl/airpollution/no2.php.

Editor Comments:

The manuscript reports on a study aimed at improving NOx emission estimates for the UK. While this is an interesting and relevant topic, a number of problems have been identified in the public discussion, including

- the use of cloudy satellite data (qa > 0.5) which is not recommended

This has been addressed in our response to M. Pommier's comment number 5. We now use QA flag > 0.75.

- the use of an outdated emission inventory

This has been addressed in our response to M. Pommier's comment number 1. We now use 2019 emission data.

- conceptual problems such as the unclear justification for the background correction method

This has been addressed in our response to Reviewer 1's 3.3 comment. We now provide justification for this.

- interpretation problems such as the uncertainties introduced by sector dependent emission trends

As we have now updated the NAEI emissions to use the most recent version for 2019 and no longer use GEOS-Chem simulations, this should resolve this issue/comment.

Considering these points, I think that major data analysis changes and text revisions would be necessary before a revised version of this manuscript could be considered for publication in ACP. Unless the authors can show evidence for actual improvements in the UK bottom-up NO_x emission estimates achieved by this work, the revisions will also have to include a change in title.

We have now provided a detailed and comprehensive response to the useful and valuable comments from the two reviewers' comments, M. Pommier and the editor. The resulting changes to the text will lead to an improved, clearer and more focussed paper. Therefore, we feel that the revised manuscript should be suitable for publication in ACP. We have also updated the title of the manuscript (see response to Reviewer 1's 3.1 comment) to "Exploiting satellite measurements to explore uncertainties in UK bottom-up NO_x emission estimates".

Additional References:

Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weißenmayer, A., Richter, A., DeBeek, R., Hoogen, R., Bramstedt, K., Eichmann, K., Eisinger, M. and Perner, D.: The Global Ozone Monitoring Experiment (GOME): Mission Concept and First Scientific Results, *Journal of Atmospheric Sciences*, 151-175, doi:10.1175/1520-0469(1999)056<0151:TGOMEG>2.0.CO;2, 1999.

CAMS. CAMS global reanalysis (EAC4). Available: https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-reanalysis-eac4?tab=form (accessed 29th November 2021), 2021.

ECMWF. ERA5 hourly data on pressure levels from 1979 to present. Available at: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview (accessed 29th November 2021), 2021.

Pena-Angulo, D., Reig-Gracia, F., Dominguez-Castro, F., Revuelto, J., Aguilar, E., van der Schrier, G. and Vicente-Serrano, S. M.: ECTACI: European Climatology and Trends Atlas of Climate Indices (1979-2017), *Journal of Geophysical Research: Atmospheres*, 125, e2020JD032798, doi: 10.1029/2020JD032798, 2020.

Torres, O., Bhartia, P. K., Jethva, H. and Ahn, C.: Impact of the ozone monitoring instrument row anomaly on the long-term record of aerosol products, *Atmospheric Measurement Techniques*, 11, 2701-2715, doi:10.5194/amt-11-2701-2018, 2018.

Tsagatakis, I., Richardson, J., Evangelides, C., Pizzolato, M., Pearson, B., Passant, N., Pommier, M. and Otto, A.: UK Spatial Emissions Methodology: A report of the National Atmospheric Emission

Inventory 2019. Available at: https://uk-air.defra.gov.uk/assets/documents/reports/cat09/2107291052_UK_Spatial_Emissions_Methodology_for_NAEI_2019_v1.pdf (accessed 29th November 2021), 2021.

Vohra, K., Marais, E. A., Suckra, S., Kramer, L., Bloss, W. J., Sahu, R., Gaur, A., Tripathi, S. N., Van Damme, M., Clarisse, L. and Coheur, P., F.: Long-term trends in air quality in major cities in the UK and India: a view from space, *Atmospheric Chemistry and Physics*, 21, 6275-6296, doi:10.5194/acp-21-6275-2021, 2021.