Author Responses to Reviewer Comments

We thank the reviewers and the Editor for their useful comments and feedback. We are pleased to see that Reviewer 2 considers that the manuscript can be accepted as it is. For Reviewer 1 and the Editor, we have reproduced their 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 re-submitted manuscript. We have also numbered the comments to make it clearer where we have provided a response.

Reviewer 1's Comments:

1. The authors have applied a simple plume model fitting to OMI and TROPOMI data to derive NOx emissions from cities in the UK. Usually this inversion method is applied to isolated sources and in this manuscript it does not become clear to me how well this works for a region as the UK which has many other sources outside the main cities.

While the reviewer is correct that this methodology has been applied to more remote sources in the past, here we are confident that are approach is suitable given the reasonable agreement between the top-down and bottom up emissions shown in **Figure 6**. There will be differences between the two emission data sets, but the strong correlation between them, and the fact that both data sources show smaller emissions for smaller sources (e.g. Edinburgh) and larger emissions for larger sources (e.g. London), provides confidence in our approach.
The fact that TROPOMI has such high spatial resolution compared with its predecessors means that we are able to isolate previously undetected sources over the UK for certain wind directions and derive emissions. It is true that for several UK cities where there are immediate large sources upwind, this approach does not work. However, where there is a smaller NO_2 enhancement over a larger spatial area (i.e. an increase in the background value due to long-range transport) upwind from sources much further upwind, as demonstrated in **Figure 3**, the running t-test is able to detect when there is a plateau in the NO_2 gradient from the source of interest (i.e. Birmingham in **Figure 3**), which is superimposed on top of the large-scale background enhancement.

2. The definition of background is not well defined and seasonal effects in the background are not discussed. I also miss some sensitivity studies in this paper about the effects of certain assumptions/choices in the method (related to lifetime, source size, and background) on the resulting emissions to have a better estimate of the uncertainty. I have several other concerns listed below and therefore I recommend to make major revisions before publication in AMT.

Based on the previous reviews in ACPD we have already provided a response and updated the manuscript in relation to seasonality of the emissions. In our previous response document this was in relation to comment 3.2c from Reviewer 1. As a result, we had added new lines to the original manuscript (now lines 279-290 in the resubmitted version). In terms of the lifetime, this is a function of wind speed and the e-folding distance. The uncertainty in the lifetime has now been incorporated into the total error term. We have updated the text in the relevant places (e.g. lines 259-262). In terms of wind speed, a key component when deriving the lifetime, in the previous round of reviews we added Figure 1, which explores the sensitivity of the wind speed, and thus lifetime, on emissions based on the altitude used. In terms of the source size, as stated on line 224, the width is a subjective choice based on the where the edges of the source are. However, for consistency, the same region is used for both the top-down and bottom up emissions. We do agree, though, that this could be made slightly clearer by mentioning the length of the source. Therefore, we have modified line 224 from “Though the source width is a subjective choice ” to “Though the source width and length are subjective choices”. As for the background derivation, we have elaborated on this in our response to Reviewer 1’s comment 4.

3. Equation 1 is introduced here without much explanation on the geometrical situation. An example of such a plume, the satellite data and the gridded data would help here. Also a reference to similar methods would help the reader.

Figure 2 has been designed to give an example of the plume e.g. for London under westerly flow so provides an example of a plume, the satellite data used on the grid. However, in line with the Editor’s final comments (4), we have updated **Figure 2a/3a** to show the line density we have used. In terms of methodology, we already make it clear on lines 108-110 that we are using a similar methodology to that of Verstraeten et al. (2018) (i.e. simple mass balance).

4. Are the grid boxes always defined along the direction of the wind or is the grid box length depending on the wind direction? Most important question here is: how is the background B defined? The background depends on sources in the neighborhood (villages, sol emissions, factories, traffic) and transport of NO2 on higher altitudes in the troposphere. Those various contributions causes the background to have a large spatially variability.

The TROPOMI TCNO$_2$ has been mapped onto a 0.025° × 0.025° horizontal resolution. The downwind transect of the data is then selected in the relevant wind direction. The downwind NO$_2$ line density is then plotted as a function of distance downwind (as shown in **Figures 2** & **3**). The background, as in other studies that have used this approach (e.g. Beirle et al., 2011 and Verstraeten et al., 2018), is derived using the downwind NO$_2$ line density. In Beirle et al., (2011) and Verstraeten et al., (2018), the background is one of the fit parameters. However, in our study the running t-test is used to determine where there has been a plateau in the line density (i.e. reached background). This is shown in **Figure 2** where there is a tail-off of the plume from London, reaching background levels before features over continental Europe are observed. The fitting of the background using the approach of Beirle et al. (2011) would struggle to resolve an accurate background value in this instance.

However, to try and make this clearer, we have reworded lines 242-253:

"Therefore, to determine when *B* has been reached, a running t-test was applied to the wind-flow *NO2 LD* profile to determine where turning points or levelling off occurred. As such a test can be sensitive to noise in the TCNO2 data, a 10-pixel (0.5°) running average wind-flow *NO2 LD* profile was calculated. The running t-test was applied to this using two windows of the same size to identify step changes in the profile. The green line in **Figure 2b** shows where the t-test p-value has become large and there is a turning point in the wind-flow *NO2 LD* profile. Such a reduction in the wind-flow *NO2 LD* profile gradient is suggestive of the plume reaching *B* as NO2 levels have stabilised. However, in **Figure 2b**, there are multiple locations potentially meeting this criteria. In reality, the turning points further downwind of London are sources from the Benelux region. The red dot represents the first instance, after the initial near-source wind-flow *NO2 LD* peak, where the gradient in the running t-test p-value profile changes sign (i.e. positive to negative or vice versa)."

to the following:

"Therefore, to determine when *B*, in the downwind direction, has been reached, a running t-test was applied to the wind-flow *NO_2 LD* profile to determine where turning points or levelling off occurred. Such a substantial change in the *NO_2 LD* profile gradient is indicative of the background level being reached and potentially another source being identified (e.g. in **Figure 2b** there is evidence of other NO_2 sources downwind of London several hundred kilometres away over continental Europe). As such a test can be sensitive to noise in the $TCNO_2$ data, a 10-pixel (0.5°) running average wind-flow *NO2 LD* profile was calculated. This smoothed out the noise from the downwind profile and allowed for the detection of larger-scale *NO_2 LD* changes. The running t-test was applied to this using two windows (i.e. a moving centre point with a window each side of 0.5°) and the t-test significance between the two window averages determined. This yielded a t-test significance/p-value distance series from the source. When a substantial change in the *NO_2 LD* gradient occurred, the t-test p-values values would increase, peak and then drop off. This change in the gradient of the t-test p-values identified the location of any *NO_2 LD* step changes in the profile. The green line in **Figure 2b** shows where the t-test p-values peaked and that there are turning points in the wind-flow *NO2 LD* profile. Such a reduction in the wind-flow *NO_2 LD* profile gradient is suggestive of the plume reaching *B* as NO_2 levels have stabilised. However, in **Figure 2b**, there are multiple locations potentially meeting this criteria. In reality, the turning points further downwind of London are sources from the Benelux region. The red dot represents the first instance, after the initial near-source wind-flow *NO2 LD* peak, where the gradient in the running t-test p-value profile changes sign (i.e. positive to negative or vice versa)."

Line 176: How is a distance between source (which has an extension of a whole city) and the background (which is everywhere) defined?

The distance between the source and background region is the distance between the source centre and the point B in the line density function. As shown by other studies (e.g. Beirle et al., 2011), this is a common approach to derive the background value in the downwind plume.

Line 185: Why should the wind speed be higher than 2 m/s for this method? And is there also a maximum wind speed? A maximum wind speed seems important to avoid situations with very long plumes covering other cities.

Here we have followed the approaches of Beirle et al. (2011) and Verstraeten et al. (2018). Both studies used a minimum cut-off wind speed but did not include an upper wind speed threshold. Beirle et al., (2011) use a minimum cut-off of 2 m/s and Verstraeten et al. (2018) use a minimum cut off of 5 m/s. Here, we have used 2 m/s like Beirle et al. (2011) as it means we get a larger frequency of samples in our study, which we have discussed in our manuscript as the UK experiences often cloudy and rainy weather conditions.

Line 224: how is the subjective choice of the source width affecting the resulting emissions?

Please see our response to Reviewer 1's comment 2.

Line 220,236: What is point B in this context? Earlier it was defined as the background, which is everywhere.

We have provided a detailed response above to Review 1's comment 1.

Line 254-265: the difficulty with this method is that the lifetime of NOx can be very different in rural regions compared to the lifetime in big cities. Therefore the plume will no longer follow the exact exponential shape. This means that the method will work better in summer than in winter time when the plume will be longer.

As discussed in response to Reviewer 1's comment 1, the running t-test can be used to determine when a city source's plume has plateaued and reached the background value, which, as the reviewer points out, will be larger in winter. Please see our response Reviewer 1's comment 1 for a more detailed response.

LIne 275: The final emission uncertainty is defined as consisting of two components of similar magnitude,(1) from satellite errors and (2) from determining the lifetime. However, I cannot find any resulting total uncertainty in the manuscript. In the table they are mentioned separately and in the Figures only component (1) is used. It seems fair if the authors will be using the total uncertainty in all Figures and Tables.

Based on this review comment, we have now replaced the two error metrics currently used in the paper with the total uncertainty. This is based on the key inputs into the emissions equation (i.e. Equation 1). These include $TCNO_2$ data (moles/m^2), the background value B (moles/m^2), the distance between the source and B (d in metres), the source width (w in metres), the wind speed (ws, m/s), the e-folding distance (x_o in metres) and the NO_2 to NO_x conversion factor (f).

Firstly, we have modified Equation 1 to make it clearer. It is now:

$$E = \frac{\sum_{i=0}^{N}((NO_2\ LD_i - B\ LD) \times \Delta d)}{t \times e^{\frac{-t}{\tau}}} \times f$$

The corresponding text in the manuscript has been modified to (i.e. lines 172-179):

"where E is the emission rate (moles/s), $NO_2\ LD$ is the NO_2 line density (moles/m), $B\ LD$ is the background NO_2 line density value (moles/m), Δd is the grid box length (m), i is the grid box number between the source and background value, t is time (s) and $e^{\frac{-t}{\tau}}$ is the e-folding loss term with τ as the effective lifetime. N represents the number of satellite $TCNO_2$ grid boxes between the source and background level B. t is calculated as the distance between the source and B divided by the wind speed (ws). To derive the full NO_2 loading emitted from the source, the wind flow $NO_2\ LD$ has the background $NO_2\ LD$ (i.e. $B\ LD$) value subtracted from all points between the source and B and is then summed yielding the total NO_2 mass (moles). f is the factor required to convert to NO_x emissions."

and then on lines 267-270:

"The top-down E is calculated from **Equation 1** and this emissions flux of NO_2 (moles/s) is converted to emissions of NO_x (moles/s) using the factor f for comparison with the bottom-up inventories. This is done by scaling the NO_2 emissions by 1.32 based on the $NO:NO_2$ concentration ratio (0.32) in urban environments at midday (Seinfeld and Pandis, 2006; Liu et al., 2016)."

Therefore, if we rewrite Equation 1 as function of the inputs with associated errors (i.e. time *t* and the e-folding lifetime *τ* are based on *d* and x_o, divided by the wind speed) we get:

$$E = (\overline{NO_2} - B) \times ws \times w \ \times f \times e^{d/x_o}$$

where $\overline{NO_2}$ is the average NO_2 value (moles/m^2) across all grid cells between the source and the background value, *B* is the background NO_2 value and *d* is the distance between the source and *B*. We can treat $\phi = \overline{NO_2} - B$ as the average enhancement (moles/m^2) above the background and therefore, derive the total error expression (i.e. Equation 3 in the manuscript):

$$\Delta E = E\sqrt{\frac{\Delta\phi^2}{\phi^2} + \frac{\Delta ws^2}{ws^2} + \frac{\Delta w^2}{w^2} + \frac{\Delta f^2}{f^2} + \frac{d^2}{x_o^2}\left[\frac{\Delta d^2}{d^2} + \frac{\Delta x_o^2}{x_o^2}\right]} \qquad (3)$$

Here, based on Beirle et al. (2011), we have assigned an error value of 10% to *w* and *f*. As *d* and x_o are both distance metrics as well *w*, we have assigned these variables errors of 10%. For *ws* we have used the standard error in the sample (i.e. all the *ws* values for a certain wind direction at a source). In retrospect, the standard error, instead of the standard deviation, is a more representative metric for error. For ϕ, we take the largest precision error from the sample (i.e. $TCNO_2$ values between the source and *B*) as a conservative estimate. For London, under westerly flow, we determine a total error of approximately 32%.

As a result, we have added the new Equation 3 on line 291 with the following text:

"To investigate the total errors in the derived NO_x emissions from TROPOMI, we have included errors from all the input terms. These include the enhancement in the $TNCO_2$ data, the e-folding distance x_o, the wind speed *ws*, the source width *w*, the NO_2 to NO_x conversion factor *f* and the distance *d* between the source and *B.* When combined, this yields the total error in Equation 3:

$$\Delta E = E\sqrt{\frac{\Delta\phi^2}{\phi^2} + \frac{\Delta ws^2}{ws^2} + \frac{\Delta w^2}{w^2} + \frac{\Delta f^2}{f^2} + \frac{d^2}{x_o^2}\left[\frac{\Delta d^2}{d^2} + \frac{\Delta x_o^2}{x_o^2}\right]} \qquad (3)$$

In the total error expression, we have set $\phi = \overline{NO_2} - B$, where $\overline{NO_2}$ is the average $TCNO_2$ value (moles/m^2) for all grid cells between the source and *B* (i.e. background $TCNO_2$ value) in the downwind profile. Here, we take $\phi \times d \times w$ to be a suitable estimate of the full NO_2 emission loading from the source (i.e. the numerator of Equation 1). Regarding the errors (i.e. terms with Δ in front), based on Beirle et al., (2011), we assign errors of 10% to *f* and *w*. As x_o and *d* are distance metrics as well, with no clear way to quantify the errors in these terms, we have assigned them with 10% errors also. The *ws* error is based on the standard error in the sample (i.e. the number days selected for each flow regime). For the enhancement in $TCNO_2$ from the source (i.e. ϕ), we have conservatively taken the largest precision error value from all $TCNO_2$ values between the source and B, which forms $\overline{NO_2}$.".

We have updated the error term values in Figures 2, 3 & 6 and in Table 1 accordingly.

Line 283: Since February and March 2018 of the TROPOMI NO2 data set contains many gaps, it is generally recommended not to use them.

The reviewer makes a relevant point here, but out of the 59 days at the start of the mission, approximately 40 days have data over the UK and Europe (i.e. ~68% of data daily data coverage). Therefore, as the UK is subject to cloudy and wet weather, limiting the retrieval of NO_2, it seems to be a bit counter intuitive to lose 40 days of data. Secondly, as the COVID-19 pandemic hit in early 2020, this severely restrictions use of the data beyond this point as it will not be representative of normal emission conditions. Therefore, for our study, it makes sense to retain the February and March 2018 data.

5. As I understand this the results are based on daily results averaged over the whole time period. To get a feeling of the quality of the results I would like to know how many days/samples are in used in the in the averaged results and what is the distribution of the samples? It can also be expected that there are differences between summer and winter results. In the summer lifetime is shorter, but on the other hand the emission will be affected by soil emissions from the agricultural regions in the UK.

In terms of the seasonality on lifetimes and emissions, we have discussed this in response to Reviewer 1's comment 2. However, we have now added the number of days used to derive the emission for each source/wind direction in each column of Table 1.

Line 202: "based on"

This has been corrected.

Line 218: equation 2: w is not defined.
This already defined on line 223 "and *w* is the source width (m)".

Line 298: I would call it "clear" instead of "sharp". the gradients can still be hundreds of kilometres.

This has been corrected.

Editor Comments:

As you will see, the two reviews are again quite contradicting. The third reviewer raised some points which you should clarify in the text. The main issue as I see it is a lack of clarity with respect to the background treatment, and I have some additional suggestions / questions on this.

1. Please explain how the width over which the line densities are integrated is determined.

As suggested in our response to Reviewer 1's comment 2, the width of the source is based on a subjective choice of the lons and lats representing the extent of the urban area of the source. Based on these positions, the distance between them is calculated and represents the source width. In terms of the line density, the $TCNO_2$ rows downwind of the source are averaged together (i.e. all the $TCNO_2$ grid rows in the downwind direction are averaged in the width direction between the edges of the source) across the width of the city. This is outlined in Equation 2 and corresponding text. However, to make this clearer, on line 216 after "source-width-average $TCNO_2$ profile" we have added "(i.e. for each downwind grid box from the source, the corresponding perpendicular rows between the source edges are averaged together)".

2. Please indicate, how mean wind speeds were determined. Were only wind speed values averaged, for which TROPOMI measurements are used?

Yes, this is correct. On line 186 we have added "Wind data is only used on days where there is TROPOMI NO_2 data available downwind of the target source, when deriving the average directional wind speed.".

3. The way I understand your approach, you subtract the NO2 line density at the point B as background, resulting in a value of 0 at point B. Is that correct? And if so, how can this be brought into agreement with the assumption of an exponential decay of NO2 (which would continue after B, leading to negative values of NO2 in the absence of fresh emissions?).

The editor is correct that for all NO_2 line density values between the source and B, the NO_2 line density value is subtracted from them essentially yielding B with a value zero. However, the fitting of the exponential decay is done before this step and the assumption is that the fitting is only valid between the peak NO_2 value and B. For other studies (e.g. Beirle et al., 2011), the fitting involves a convolution of a Gaussian distribution and an exponential decay function. Here, the exponential decay function describes the decay of the plume and the lifetime determined. Therefore, our approach is suitable between the peak of the plume and B. This is inferred on lines 254-255 “The loss term $e^{-t/\tau}$ is dependent upon τ and is determined by applying an e-folding distance fit between the near-source peak wind-flow *NO2 LD* value and B, before dividing by ws to get τ.”, but after the first mention of B we have added “(i.e. we assume this function is valid only between these two points)” to make this clear.

4. In Figure 2a, it would be very helpful to add a box indicating the region over which the line density was determined. Also in Figure 2c, it would be good to show point B and the background which you subtract from the line density. I'd also suggest to change the y-axis of this plot to start at 0.

For Figures 2a and 3a we have added a box to show where the line density is that we are using.