Author Responses to Reviewer Comments

We thank the reviewers for their useful comments and feedback. 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 originally submitted manuscript.

Reviewer 1's Comments:

General comments:

1. The study uses "novel" satellite observations of tropospheric ozone from the GOME-2 and IASI instruments, but the manuscript has few details about the retrievals. While the GOME-2 retrieval have been described in Miles et al., 2015, a few more details about it in the manuscript would be useful. I could not find details of the IASI ozone retrievals in the cited manuscripts (Pope et al., 2021, Palmer et al., 2022). It is important to discuss the vertical sensitivity of the retrievals (averaging kernel profiles), its the vertical resolution in the troposphere, how the retrieval compares to independent (ozonesonde) observations, etc.

There is a detailed evaluation of the RAL Space IASI IMS-Extended product in the Supplementary Material of both Pope et al. (2021) and the Palmer et al. (2022) studies, rather than the main papers. We have also made reference to Pimlott et al., (2022), which includes evaluation and discussion of the IASI IMS product. Evaluation of the GOME-2 product is discussed by Miles et al., (2015), as indicated by the reviewer, and already referenced. However, to provide more information on the vertical sensitivity we have included a new figure and section in the supplementary material. Below is the new Averaging Kernel figure (now Figure SM1) for the GOME-2 and IMS schemes for the 1st June 2017 and 2018 averaged over the European domain.

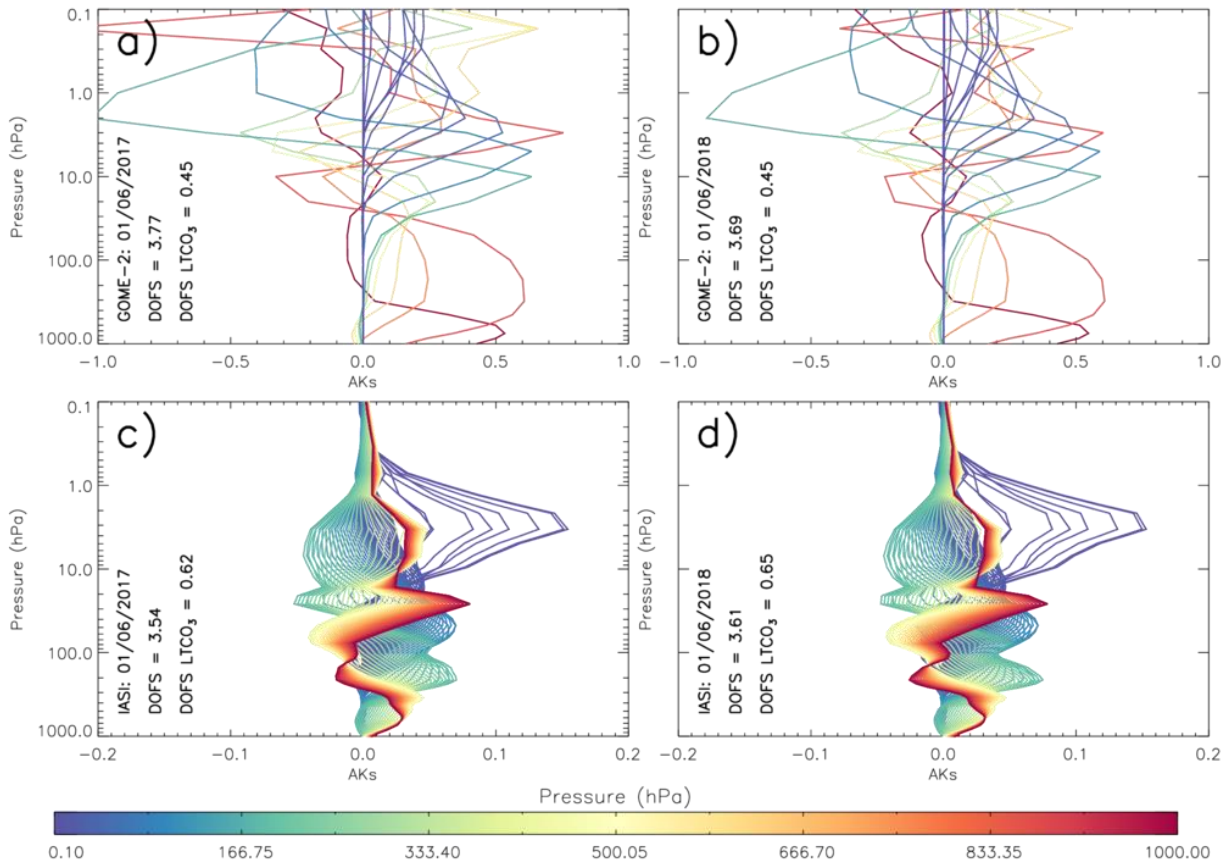

Figure SM1: Example European domain average averaging kernels (AKs) for retrieved ozone profiles for a) GOME-2 on the 1st June 2017, b) GOME-2 on the 1st June 2018, c) IMS on the 1st June 2017 and d) IMS on the 1st June 2018. The degrees of freedom of signal (DOFS) for the full profile and lowest sub-column ($LTCO_3$, surface – 450 hPa) are also shown.

New Supplementary Material (SM) 1 section:

"**Supplementary Material (SM) 1 – Satellite Ozone Averaging Kernels**

This work has used satellite lower tropospheric column ozone ($LTCO_3$, surface – 450 hPa) data produced by RAL from the Global Ozone Monitoring Experiment – 2 (GOME-2) and by their extended Infrared and Microwave Sounding (IMS) scheme applied to IASI, MHS and AMSU on MetOp-A. For each of these products, **Figure SM1** shows averaging kernels (AKs) for the 1st June 2017 and 2018 averaged over the European domain. The AK shows the sensitivity of the retrieved profile at different levels to a perturbation in the true profile at different levels (Rodgers, 2000; Eskes and Boersma, 2003). Therefore, it represents the instrument's vertical sensitivity to retrieving ozone. For both GOME-2 and IMS, AK shapes show that retrieval levels at lower altitudes have sensitivity to troposphere ozone. The degree of freedom of signal (DOFS) for $LTCO_3$ for GOME-2 is 0.45 while IMS ranges between 0.62 and 0.65. Therefore, there is good information in these products on $LTCO_3$, which remains similar between the two years, so we are justified in using these $LTCO_3$ products in this study."

2. The model used for investigating the ozone pollution episodes in 2018 does not seem well-suited for this application. In particular, it fails to capture the observed increase in ozone in 2018 compared to 2017, suggesting that the sensitivity of the model to emission or meteorological charges is too low. Also, a 2.8°x2.8° horizontal resolution seems too coarse for regional modeling.

The reviewer is correct that using a higher resolution model would be preferable. However, the aim of this study is to investigate the broad scale tropospheric ozone increases in the summer 2018. To do this, we need to use a global model with transport of ozone from outside the European domain (both stratospheric and tropospheric sources). Using a regional model makes this more difficult for several reasons. 1) The boundary conditions are normally from model simulations using more coarse spatial resolutions (e.g. 4 or 5° longitude-latitude grid) than used here. 2) Regional setups either have poor or no representation of the UTLS. 3) They are extremely time consuming to run, so investigation of processes through sensitivity experiments becomes much more difficult (e.g. the fixed met, fixed emission, fixed deposition experiments – we did undertake several other experiments as well which were not discussed here – could not be done realistically in a regional or nested model). Therefore, while TOMCAT does not simulate the full magnitude of the 2018 ozone response seen by the observations or individual city plumes leading to secondary formation, it is still a very useful modelling framework to investigate processes influencing the broad scale distribution of tropospheric ozone. It is also worth pointing out that TOMCAT has also been used in regional studies before which have been published in peer-review journals (e.g. Richards et al., 2013; Pope et al., 2018; Pope et al., 2020). These references have been added in the discussion of TOMCAT in Section 2.2.

3. The back trajectory analysis to determine changes in ozone transport is confusing and unconvincing. The back trajectories are weighted by the modeled ozone to determine the amount of advected ozone. However, the frequency with which air arriving at a receptor site passes a particular location does not seem to be accounted for. Also, the choice of Paris and Berlin as receptor sites seems arbitrary. How does the back trajectory model handle convective transport from the boundary layer to the free troposphere? I wonder if it would be much easier just to derive the net influx of ozone into Europe from the 3D chemical transport model using the ozone concentrations and the wind fields.

So firstly, while we agree that Paris and Berlin are subjective choices, they are selected as they represent sites in central Europe, which had experienced the summer 2018 heatwave. This is discussed on Page 4 Lines 151-152 (i.e. *"both central locations over Europe in the region of summer-time 2018 O_3 enhancements"*). If this study were to be repeated, it might be possible to run ROTRAJ from every TOMCAT grid box centre point over Europe and undertake the same analysis. However, this would be extremely time consuming and unlikely to add more scientific benefit than the selected central Europe sites used. As for convective transport, it can resolve large-scale advection and convection but not subscale features. To make this clearer, we have updated the text on Page 4 Lines 146-148:

"*ROTRAJ is a Lagrangian atmospheric transport model driven by meteorology from the same ECMWF ERA-Interim reanalyses as used by TOMCAT.*" with

*"*ROTRAJ is a Lagrangian atmospheric transport model driven by meteorology from the same ECMWF ERA-Interim reanalyses (horizontal resolution of 1.0125°) as used by TOMCAT.*"*

And on Page 4 Line 150 we have added:

"This method accounts for large scale advection since the winds are resolved but does not resolve small scale sub-grid turbulent transport."

As for the frequency of trajectories reaching the receptor sites, this is not easy to quantify and account for in the weighted ozone value in each pixel on the grid. This is because when the average O_3 value from all the trajectory points in a grid box is calculated, it does not know where it has come from. However, from Figure S12 and Figure S13, qualitatively, there appears to be suitable coverage across the North Atlantic region. While some grid box O_3 values will have more information from a range of locations, by eye, there appears to be good coverage (i.e. Figure S12 and Figure S13) between 30-70°N (i.e. key region of interest discussed in this study). When the number of trajectory points is plotted in each grid box, the largest values are over the receptor sites (>500), which drops to ~100 on the edges of Europe and then 25-50 over the North Atlantic. This distribution is relatively consistent and does not change much between years. In **Figure Response 1**, it shows the percentage difference between years for the receptor sites at the surface and at 500 hPa. Overall, there is a peak change between years in grid box percentage occurrence of 10-20%. Further away from the receptor sites, this drops to <10%.

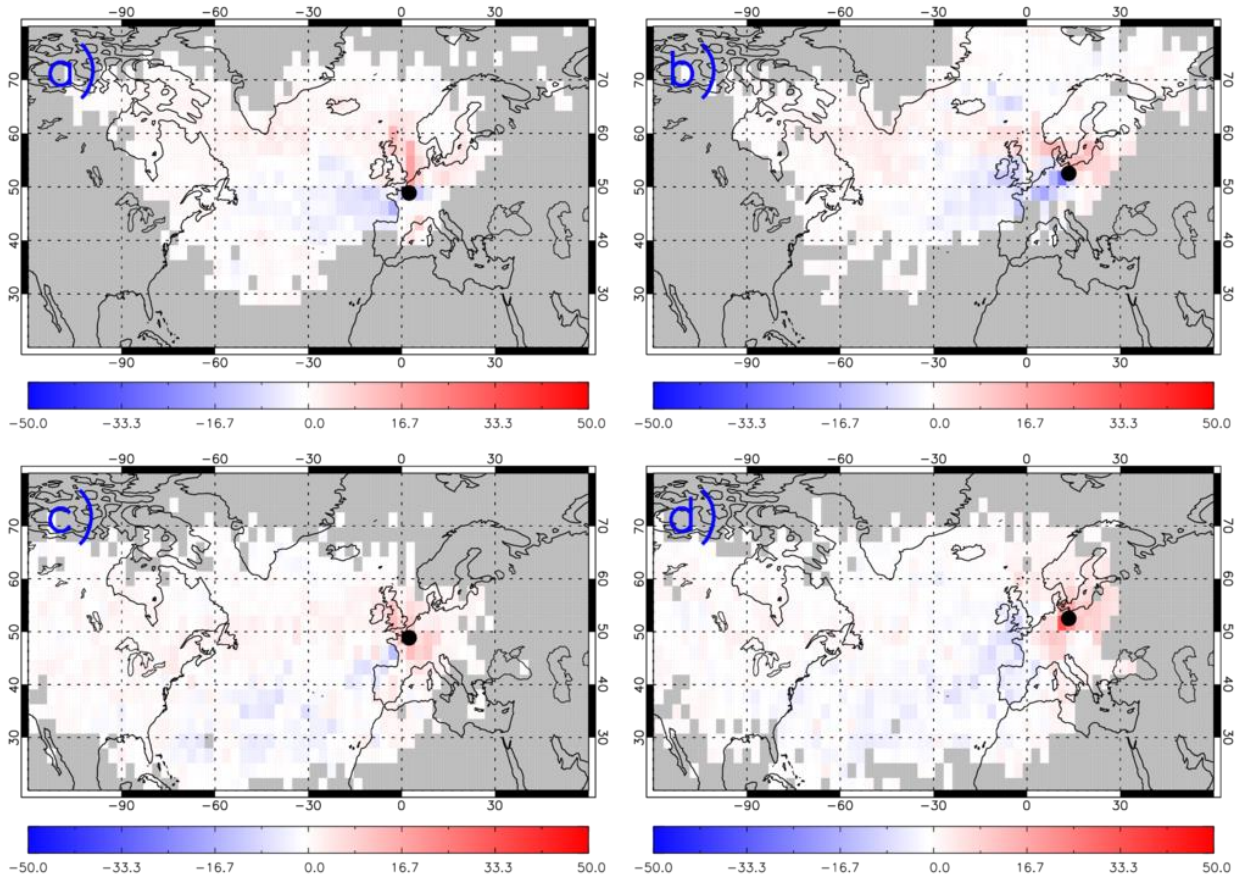

Figure R1*: The May-August 2018-2017 difference between grid box trajectory point frequencies (as a percentage) for a) Paris surface, b) Berlin surface, c) Paris 500 hPa and d) Berlin 500 hPa.*

Therefore, overall, the frequency of trajectories does not appear to be substantially different and the approach we used to weight the trajectories by TOMCAT O_3 and get gridded O_3 trajectory values (i.e. Figure 12 and Figure S14 and Figure S15) is suitable for this study. To make this clear we have added on Page 9 Line 382:

"While this approach does not directly account for the frequency of trajectory points in each grid box, **Figures SM 12** and **13** show there is wide spread coverage across the North Atlantic. This results in >500 trajectory points near the receptor sites (i.e. Paris and Berlin), ~100 trajectory points around the edge of Europe and 25-50 trajectory points in the North Atlantic (not shown here). Overall, this spatial distribution is relatively consistent and does not change substantially between years (typically 10%), thus this approach is suitable in this study."

In terms of calculating fluxes, this approach can tell you the amount of O_3 and where it is coming from for each grid cell in terms of the u- and v- wind components. However, as shown by the back trajectories, there are substantial variations in them (i.e. motion from south-north and east-west), so just looking at a map of O_3***u** or O_3***v**, or a cross section at a particular latitude or longitude (e.g. at 0°E, 30-70°N), will not easily tell you the origin of the O_3. Hence, this is where the trajectory method has its benefits. It is also worth pointing out, this sort of approach with back trajectories, has been done before (e.g. Stirling et al., 2020 and Graham et al., 2020). Thus, we have added on Page 4 Line 161:

"This follows a similar approach to Graham et al., (2020) and Stirling et al., (2020), though using a model chemical tracer and not emission inventories.". These references have been added to the reference list.

4. Could the authors include in a discussion of what the results of this work imply for our understanding of ozone air pollution during heat waves? The results, as presented currently, seem to apply only to the particular episode in the summer of 2018, which limits the scientific significance of the work. Do any of the results apply more broadly?

The general focus of this study has been on the summer 2018 heatwave and several of the features (e.g. the enhancement in stratospheric ozone in the mid-troposphere) are likely specific to this event. However, this does not necessarily mean the results are less significant or important. The tools used (e.g. novel satellite data products with good vertical sensitivity to lower tropospheric ozone and the targeted model sensitivity experiments) represent new and substantial work. They represent a data-modelling framework which other users can adopt for future studies. Therefore, we believe our study is still highly relevant. To make this point clearer, we have added on Page 11 Line 465:

"Overall, through our study has focused on the European summer 2018 air pollution episode, we have demonstrated the use of novel satellite datasets and a modelling framework (i.e. targeted sensitivity experiments and model tracers) suitable to investigate the air quality impacts from future European heatwaves such as that which occurred in summer 2022."

Specific comments:

1. The introduction lacks a discussion of previous work on this topic.

Please see our response to review #2's first Minor Comment.

2. Line 100: “cost function less than 200” – This needs some context.

One Page 3 Line 102, “a cost function less than 1000.0” is incorrect and should be <200 as stated on Page 3 Line 100. Thus, probably the confusion and the former has been removed. However, even if not, we provide the quality control flags used, so if other users of the data wish to investigate our results, they know the quality assurance flags we have used. The full details of the different variables are contained in Miles et al., (2015), which is referenced on Page 3 Line 89. The purpose of this study is to exploit the data for scientific investigation rather than recap on details of the satellite data auxiliary variables.

3. Line 104: How were the 83 sites selected out of the hundreds of sites in the EMEP network?

When this work was initially undertaken these were the number of sites available with data for both years. This has now been extended to use the new data which has been uploaded to the EMEP observational site since. There are now 125 temporally co-located sites from EMEP used in the updated Figure 5 of the manuscript and Figure S7 of the supplementary material. These two figures are shown below:

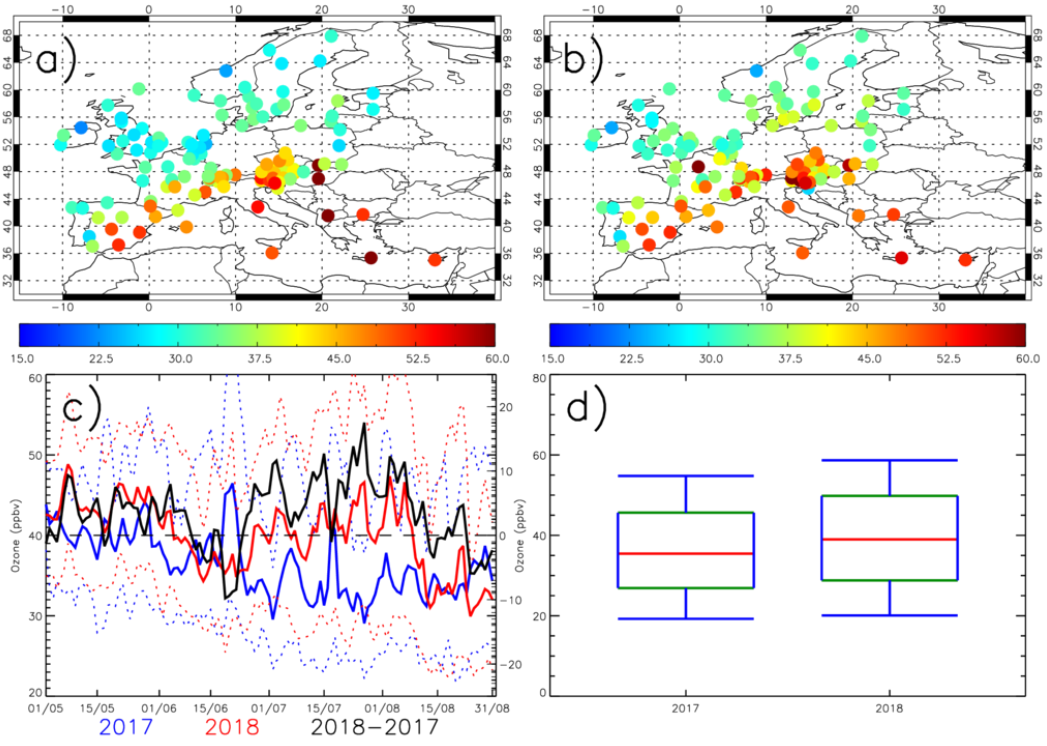

Figure 5: *European surface ozone (ppbv) for a) May-June-July-August (MJJA) 2017, b) MJJA 2018), c) regional mean time series (dotted lines show mean ± standard deviation) for MJJA 2017 (blue), MJJA 2018 (red) and the 2018-2017 difference (black) and d) box-whisker plots for MJJA 2017 and 2018. In panel d) the median, 25th & 75th percentiles and 10th& 90th percentiles are shown by the red, green and blue lines, respectively.*

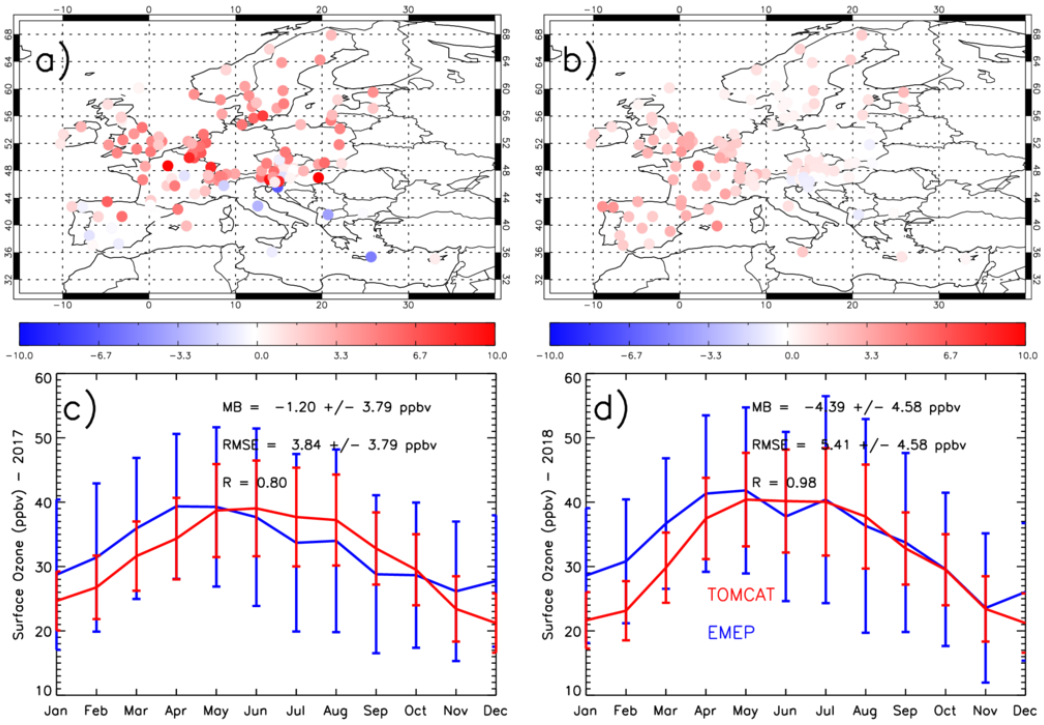

Figure S7: *Surface O_3 (ppbv) May-June-July-August 2018-2017 average difference for a) EMEP sites and b) TOMCAT. The observational (blue) and modelled (red) surface O_3 seasonal cycle are for c) 2017 and d) 2018. Vertical bars represent the monthly standard deviations. The statistics in the top right of c) and d) are the mean bias (MB), the root mean square error (RMSE) and correlation (R). The uncertainties on the MB and RMSE are the standard errors corrected for temporal autocorrelation.*

The corresponding results text has been updated accordingly to represent the updated figures.

4. Line 129: Please clarify that the fixed emission run includes fixed biogenic VOC emissions.

Yes, all the emission are fixed in the fixed emissions simulation. We have made this clearer on Page 4 Line 130 by replacing "(i.e. 2017 emissions for both years)" with "(i.e. 2017 emissions for both years – including BVOC emissions)".

5. How do biogenic and lightning NOx emissions differ between the two years?

As discussed in section 2.2, the biogenic emission are derived using the JULES model. As this is driven by meteorological reanalysis data for 2017 and 2018, there are differences in the BVOC emission responses. This is already discussed in more detail in the supplementary material SM3 and in Figure S6. As for the lightning emissions, these are parameterised from the convection scheme in TOMCAT which is based on Stockwell et al., (1999). However, to make this clearer the following sentence has been added on Page 3 Line 122:

"Lightning emissions of NO_x are coupled to convection in the model, which is derived from the meteorological reanalyses. Therefore, they vary in space and time according to the seasonality and spatial pattern of convective activity (Stockwell et al.,1999)."

6. Line 211: Please verify whether GOME-2 actually has peak sensitivity in the lower troposphere. Miles et al. 2015, which describes the GOME-2 retrieval, states that it peaks at 500 hPa. It would be useful to include the averaging kernel profiles for GOME-2 and IASI retrievals in the supplement.

The GOME-2 AKs for sub-column layers used in this analysis peak in the in the surface to 450 hPa layer. We have added a plot to the supplementary material of the GOME-2 and IASI AKs to show this and in response to reviewer #1's comment #1.

7. Line 242: Do you see any spatial consistency between the charges observed at the surface and from the satellite data?

When it comes to surface measurements and satellite retrievals of ozone, they represent very different, complementary quantities (i.e. surface ozone mixing ratio vs. a sub-column amount for which vertical sensitivity is comparatively low at surface level). Therefore, it is difficult to compare the two directly. Their spatial resolutions also differ substantially: whereas surface sites provide a comparatively sparse set of point measurements, satellite data for individual scenes of 40 x 80 km were gridded to the TOMCAT horizontal resolution. Therefore, close correspondence in spatial structure in 2018-17 difference plots is therefore not anticipated and comparisons over an average region for the surface sites is more informative (e.g. as done in Figure 5 to get a domain response in surface ozone in 2018 relative to 2017).

8. Line 248: Again, I do not think that the tendency of the model to underestimate ozone charges between 2018 and 2017 can be brushed aside.

We refer the review to our response to reviewer #1's comment #2.

9. Section 3.4 can be considerably shortened. Much of text just repeats what is shown in the figures. Better to emphasize the main takeaways instead.

We have tried to make the manuscript more concise but as there is a lot of information in Figures 6-12, it is important to add sufficient discussion on the results. These are then summarised in section 4. With the comment *"Much of text just repeats what is shown in the figures"*, this is quite general so it is difficult to make more concise changes in line with what the reviewer is asking for.

10. Line 455: The influence of stratospheric ozone intrusions is shown to be higher in 2018 than in 2017. Is this a coincidence or are more stratospheric intrusions linked to meteorological processes associated with heat waves? Also, this does not seem to be supported by the back trajectory analysis, which do not show any trajectories originating from the stratosphere.

For the back trajectory analysis, we have determined the tropopause level based on the WMO definition of "the lowest level at which the temperature lapse rate decreases to 2 K/km or less" at each trajectory point. The average tropopause pressure at each time step of the back trajectories (initialised at 500 hPa) was calculated and added to **Figures S18** & **S19**. The dashed black line represents the mean tropopause value across the trajectory sample and the dotted black lines are the mean ± standard deviation. As can be seen from the updated trajectory plots below, some of the trajectories are indeed originating from the stratosphere or are very close the average tropopause pressure. At each time step, the tropopause

pressure will vary substantially and the air mass only has to be tagged once to be labelled as stratospheric air. Secondly, the trajectories only go back 10-days, so stratosphere-troposphere exchange (STE) will also be occurring in timestamps before the final back trajectory points have been determined by ROTRAJ. In terms of the meteorological processes controlling the STE, we are primarily interested in the tropospheric processes controlling the enhanced summer 2018. Therefore, while the reviewer makes a good point "*Is this a coincidence or are more stratospheric intrusions linked to meteorological processes associated with heat waves?*", it is realistically beyond the scope of this study to quantify this with a detailed assessment of stratospheric processes controlling STE in the case of this heatwave and better suited for a separate study. So, in addition to the updated Figures S18 & S19, we have added the following two text additions on Page 9 Line 356 and Page 6 Line 242 of the supporting information:

"However, a separate study would be required to undertake a detailed assessment of the meteorological processes controlling the enhanced stratospheric intrusion of ozone in the summer of 2018 and how it compares to other years (how does it compare with other years other than 2017)."

"(i.e. the back trajectories have close proximity to the tropopause pressure – see **Figures S18** & **S19**)"

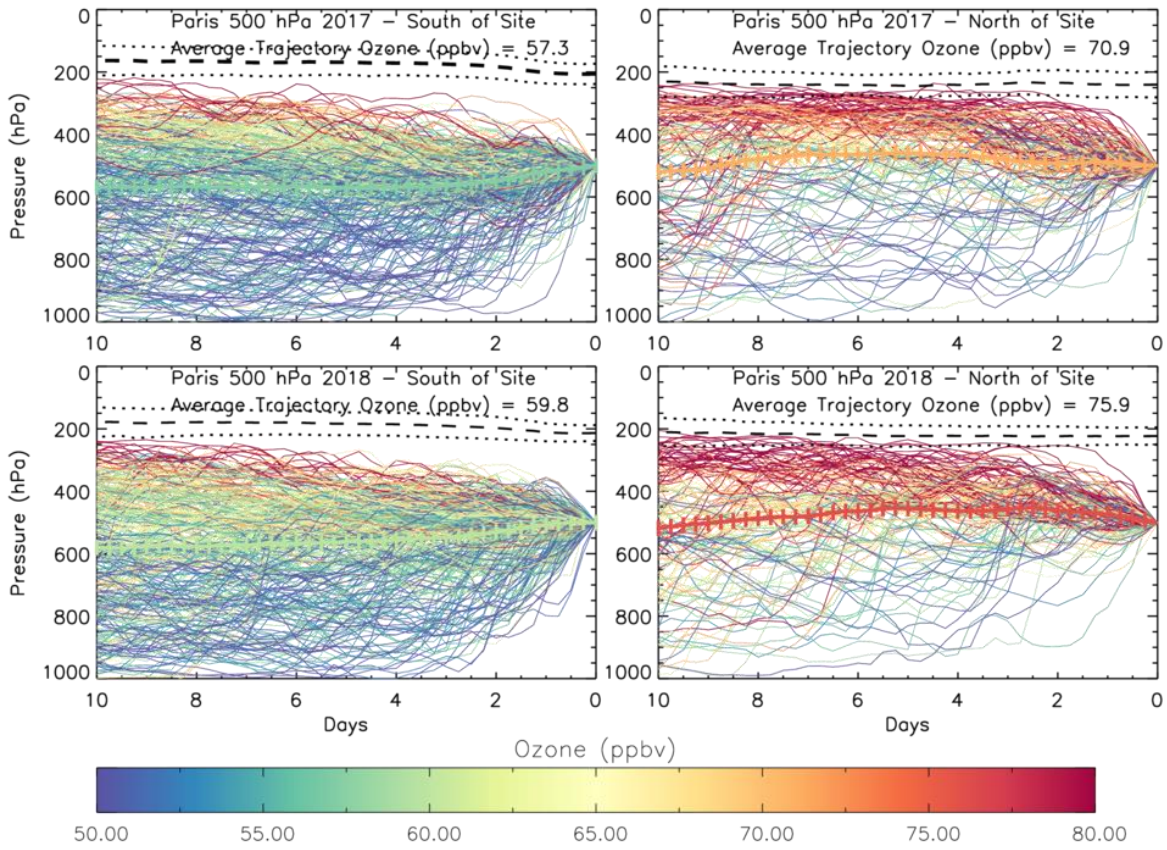

***Figure S18**: ROTRAJ back-trajectories (10 days), weighted by the average TOMCAT O_3 (ppbv) concentration along each trajectory path, plotted as time-pressure profiles, released from Paris at 500 hPa for top-left) 2017 originating south of the release point, top-right) 2017 originating north of the release point, bottom-left) 2018 originating south of the release point and bottom-right) 2018*

originating north of the release point. The thick cross lines show the average time-pressure profile coloured by the average weighted TOMCAT O_3 value. The dashed black line presents the average tropopause pressure at each time-step of all the trajectories. The dotted black lines show the average tropopause pressure ± the standard deviation at each time step.

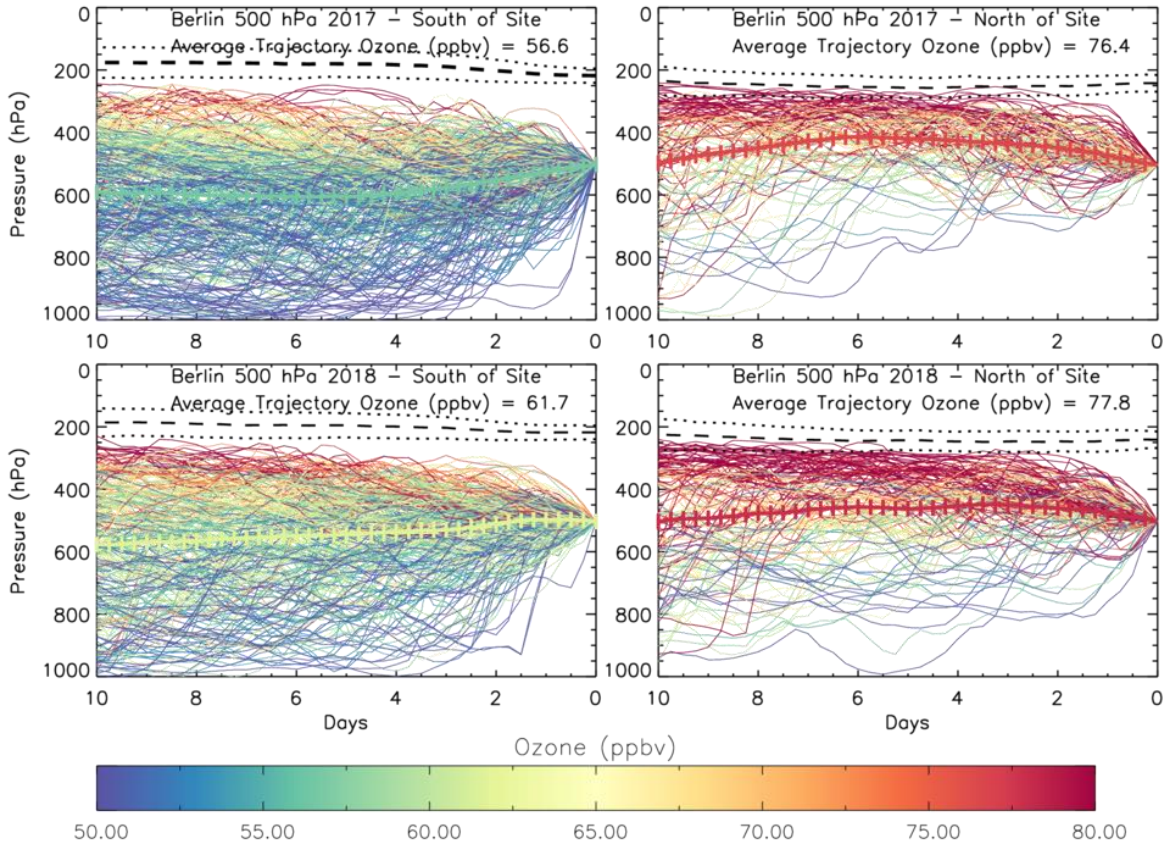

***Figure S19**: ROTRAJ back-trajectories (10 days), weighted by the average TOMCAT O_3 (ppbv) concentration along each trajectory path, plotted as time-pressure profiles, released from Berlin at 500 hPa for top-left) 2017 originating south of the release point, top-right) 2017 originating north of the release point, bottom-left) 2018 originating south of the release point and bottom-right) 2018 originating north of the release point. The thick cross lines show the average time-pressure profile coloured by the average weighted TOMCAT O_3 value. The dashed black line presents the average tropopause pressure at each time-step of all the trajectories. The dotted black lines show the average tropopause pressure ± the standard deviation at each time step.*

11. Figure 5: What do the dotted lines show?

These show the mean value ± the standard deviation in the time-series. This has been made clearer in the caption of Figure 5.

12. Figure 5: Wouldn't MDA8 ozone be a better metric for this?

We agree with the reviewer on this but the model data has been output at 6-hourly interviews primarily to compare with the satellite data. Therefore, while we could have derived the MDA8 metric from the observations, we could not do so for the model. Also, as we are not trying to estimate the health burden from the ozone pollution, it is not critical that we present this metric. Therefore, we believe the current metric used is suitable in our study.

Reviewer 2's Comments:

Major comments:

The model horizontal resolution is fairly coarse and it's likely to average out any large regional emissions of ozone precursors. As well as the model representation of ozone precursors emissions, the coarse model resolution is likely to impact on ozone formation and ozone deposition; these impacts will be largest near the surface. Can the authors justify the choice of model resolution for this study, particularly when looking at surface ozone and comparing to EMEP surface station data? Similarly, is the vertical resolution in this model appropriate for this study? The authors should give an idea of the vertical resolution, e.g. number of levels in the boundary layer and mid latitude free troposphere, in section 2.1.

Please see our response to Reviewer #1, General Comment #2 for the model resolution comment from Review #2. In terms of the vertical resolution, Figure 1 of Monks et al., (2017) shows that there are approximately 6 model levels in the boundary layer (800 hPa) and approximately 10 model levels in mid-troposphere (800-400 hPa). We have updated the text on Page 3 Lines 109-110 "*In this study the TOMCAT CTM (Chipperfield, 2006) is forced by European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis meteorology (Dee et al., 2011) and run at a horizontal resolution of 2.8° × 2.8° with 31 vertical levels from the surface to 10 hPa.*" to "In this study the TOMCAT CTM (Chipperfield, 2006) is forced by European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis meteorology (Dee et al., 2011) and is run at a horizontal resolution of 2.8° × 2.8°. The model has with 31 vertical levels from the surface to 10 hPa with 5-7 (approximately 10) levels in the boundary layer (mid-troposphere), depending on latitude.".

The authors should consider differences in biomass burning emissions of ozone precursors between 2018 and 2017.

The fire emissions from GFAS are for 2017 and 2018 for the respective TOMCAT simulations. This has been clarified on Page 3 Line 114:

"fire emissions from the Global Fire Assimilation System (GFAS, Kaiser et al., 2012) for 2017 and 2018."

A lot of results, including model evaluation and comparison with observations, are currently in the supplementary material. I believe a lot of the analysis and discussion in supplementary material should be transferred to the main body of the paper.

Reviewer #1 has suggested that section 3.4 is too long as it is (see Specific comment #9) and the results can be condensed. As a result, we have attempted to streamline the text but generally feel that it is important to discuss the plots as they currently are in sufficient detail. Therefore, by bringing a substantial part of the supplementary material into the main manuscript would make it much longer.

The general idea was to put the model evaluation (and more technical comparisons) into the supplementary material and then discuss the key scientific messages in the paper. Therefore, we propose to leave the manuscript as it is but are happy to defer to the Editor if they feel the structure of the manuscript and supplementary material do need changing.

Figure S7 shows that TOMCAT fails to reproduce surface ozone's seasonal cycle compared to observations and also the extent of the ozone enhancement in 2018 compared to 2017, possibly due to the model's coarse resolution. This is also supported by comparison of figure S8-S11. Are the authors justified in using TOMCAT to disentangle different processes leading to the ozone enhancement in 2018?

So, in terms of the model resolution, we politely refer the reviewer to our response to Reviewer #1, General Comment 2. In terms of the seasonal cycle, the reviewer is correct that the model struggles to simulate the larger EMEP values in Jan, Feb and March, however, from April/May onwards, they are in good agreement with close overlap in the time-series. Even with the underestimation in winter/spring, the R, MB and RMSE metrics show good agreement annually, especially in summer which is the focus of this study. We have already stated in the manuscript that the model does generally underestimate the 2018-2017 differences for surface and satellite ozone. However, the model still simulates substantial 2018 ozone enhancements with good spatial agreement in comparison to the observations. And while the model slightly underestimates this response, the model provides a useful tool to investigate the sensitivity to different processes (e.g. emissions and meteorology) which is not possible with observational data alone. Secondly, as in our response to Reviewer #1, General Comment 2, this model setup means we can easily and quickly do robust sensitivity experiments while regional models would take too long to realistically undertake multiple experiments. TOMCAT also allows us to investigate processes like UTLS ozone exchanges, which regional models would not allow. Thus, TOMCAT is a suitable modelling framework for this study.

Minor Comments

The introduction should be extended to include a broader discussion on the many factors affecting Summer time European ozone, including intercontinental transport and the impact of interannual variability in biomass burning and BVOC emissions, see for example work by Honrath et al. 2004 (Honrath et al., Regional and hemispheric impacts of anthropogenic and biomass burning emissions on summertime CO and O3 in the North Atlantic lower free troposphere, J. Geophys. Res., 109, D24310, doi:10.1029/2004JD005147)

To address this comment and review #1's comment specific comment #1, we have updated the text on Page 2 Lines 58-68 from:

"As well as meteorological and vegetation responses, enhancements in atmospheric pollutants from heatwaves can lead to a degradation in air quality (AQ). Firstly, anticyclonic conditions (i.e. atmospheric blocking) have been shown to cause the accumulation of primary air pollutants such as carbon monoxide (CO; Thomas and Devasthale, 2014), nitrogen dioxide (NO_2; Pope et al., 2014) and particulate matter (i.e. $PM_{2.5}$; Graham et al., 2020) to hazardous levels. Secondly, higher temperatures during blocking events, which can trap and accumulate existing pollutants (e.g. Pope et al., 2016), can lead to the secondary

formation of tropospheric ozone (O_3). Elevated surface O_3 is associated with adverse health impacts (Doherty et al., 2017; Jerrett et al., 2009) with ailments such as asthma, reduced lung function and disease (WHO, 2021). It also has adverse impacts on the natural biosphere (Sitch et al., 2007) and agriculture (Hollaway et al., 2012; van Dingenen et al., 2009)." with:

"As well as meteorological and vegetation responses, enhancements in atmospheric pollutants from heatwaves can lead to a degradation in air quality (AQ) across Europe. Blocking systems (anticyclonic conditions) have been shown to increase the level of air pollutions such as carbon monoxide (CO; Thomas and Devasthale, 2014), nitrogen dioxide (NO_2; Pope et al., 2014) and particulate matter (i.e. $PM_{2.5}$; Graham et al., 2020) to hazardous levels. Pope et al., (2016) focused on the 2006 UK heatwave and detected enhancements in surface O_3 through the accumulation of pollutants (i.e. atmospheric blocking) but also the higher temperatures yielding more active atmospheric chemistry (i.e. ozone formation). Papanastasiou et al., (2015) found that Greek heatwave conditions (2001-2010) typically yielded an increase in NO_2, $PM_{2.5}$ and O_3 by 14-29%, 25-38% and 12%, respectively. Rasilla et al., (2019) found that heatwaves in Madrid only moderately increased NO_2 and O_3 but significantly increased PM_{10} concentrations. However, they associated this with enhanced long-range transport of African dust and then accumulation under heatwave conditions. García-Herrera et al., (2010) provided a review of the 2003 European heatwave finding that the Alpine region had substantially elevated surface ozone levels (peaking at 417 $\mu g/m^3$ with 68% of sites from 23 countries reaching concentrations above 180 $\mu g/m^3$) when compared with the previous 12 summers. Biogenic volatile organic compound (BVOC) emissions from vegetation are known to increase under drought conditions from temperature stress (e.g. in the 2003 European heatwave; Rennenberg et al., 2006). Churkina et al., (2017) found that heatwave conditions (2006) in Berlin yielded an increase in BVOC emissions which contributed up to 12% of the surface ozone formation. Heatwaves can also trigger wildfires, which emit primary air pollutions and can form secondary gases such as surface ozone on a regional and hemispheric scale (Honrath et al., 2004). Overall, elevated surface O_3 is associated with adverse health impacts (Doherty et al., 2017; Heal et al., 2013; Jerrett et al., 2009,) with ailments such as asthma, reduced lung function and disease (WHO, 2021). It also has adverse impacts on the natural biosphere (Sitch et al., 2007) and agriculture (Hollaway et al., 2012; van Dingenen et al., 2009), in turn reducing deposition of surface ozone on vegetation."

p3, l 103: the authors should say a bit more about the EMEP measurements and add appropriate references.

We have updated the text on Page 3 Lines 103-106:

"We also use surface O_3 observations from the European Monitoring and Evaluation Programme (EMEP) network for May-August 2017 and 2018. In total, we used 83 spatial collocated EMEP sites in both years years across Europe. Here, data at individual sites were selected where the corresponding data flag was set to 0.0." with

"We also use surface O_3 observations from the European Monitoring and Evaluation Programme (EMEP) network for May-August 2017 and 2018. The EMEP network contains >100 surface measurement sites measuring information on a range of air pollutions (e.g. ozone, NO_2 and $PM_{2.5}$). EMEP surface data can be used for multiple scientific applications such as trends analysis (Yan et al., 2018) and atmospheric chemistry model evaluation (Schultz et al., 2017; Archibald et al., 2020) and is hosted by the EBAS

database infrastructure, developed by the Norwegian Institute for Air Research. In total, we used 125 spatial collocated EMEP sites in both years across Europe. Here, data at individual sites were selected where the corresponding data flag was set to 0.0."

p3, l 113-114; "year-specific anthropogenic emissions from MacCity". Historic emissions for this dataset end in 2010 but the model simulations are for 2017 and 2018. Could the authors explain how these emissions were extended to cover 2017-2018? If scenarios emissions are used, what pathway do they follow, and can we say that the emissions are year-specific if they are using a scenario?

The reviewer is correct that emission inventories are not available for present day or recent years. We believe the most up to date emissions are from CMIP6 and go up to 2014. Therefore, the emissions are based on scenarios. From the emission data set, they are based on RCP 8.5 as discussed in the Granier et al., (2011) paper. However, we have updated the text to make things clearer on Page 3 Lines 113-115:

"Anthropogenic emissions used in this study come from MACCity (Granier et al., 2011). The original dataset in Granier et al., (2011) derived emissions up to 2010. Therefore, the Representative Concentration Pathways 8.5 (RCP 8.5) were used by Granier et al., (2011) to generate emissions for later years (e.g. 2017 and 2018 as used in this study). Fire emissions came from the Global Fire Assimilation System (GFAS, Kaiser et al., 2012)."

p4, l 135: can the author explain more about the O3 tracer. Does it have a fixed lifetime, does it decay once in the troposphere? Can they cite previous work using this tracer with TOMCAT or other models?

The stratospheric ozone tracer (O_{3S}) is a separate tracer which is set equal to the model-calculated O_3 in the stratosphere. The only tropospheric source of O_{3S} is transport from the stratosphere while its sinks are via photolysis and reactions with HO_2, OH and H_2O through $O(^1D)$ produced from O_{3S} and surface deposition (Monks et al., 2017). The tracer does not have a fixed lifetime but the loss rate in the troposphere depends on the modelled local OH, HO_2, H_2O and photolysis. Any O_3 that gets into the stratosphere will be labelled as stratospheric before it returns. On Page 4, Lines 134-135, we have replaced "*TOMCAT also includes a stratospheric O_3 tracer (i.e. tags O_3 in the model which originated in the stratosphere).*" with:

"TOMCAT also includes a stratospheric O_3 tracer, a common approach to tag stratospheric O_3 (e.g. Roelofs et al., 2003; Akritidis et al., 2019), which can be used to investigate the impact of stratospheric O_3 intrusion into the troposphere. The tracer is set equal to the model-calculated O_3 in the stratosphere. The only tropospheric source of O_{3S} is transport from the stratosphere while its sinks are via photolysis, reactions with HO_2, OH and H_2O through $O(^1D)$ produced from O_{3S} and surface deposition (Monks et al., 2017). The tracer does not have a fixed lifetime but the loss rate in the troposphere depends on the modelled local OH, HO_2, H_2O and photolysis. Any O_3 that gets into the stratosphere will be labelled as stratospheric before it returns."

Figure 1: the caption in Fig 1 contains the sentence "…retrieved from MetOp-A IASI, MHS and AMSU by the IMS scheme." This should be expanded, appropriate references added and moved to section

This is already defined on Page 3 Lines 94-99. Also, in response to Review #1's General Comment #1, we have added more references for the IMS scheme.

2.1. ‘LHS’ and ‘RHS’: please replace with ‘left’ and ‘right’ (same for Fig 4, 6, 7, 8) or explain somewhere what the acronym means.

We have updated the figure captions accordingly.

Figure 2: the green polygons make it hard to understand what’s significant (it’s not clear if it’s inside/outside the contour lines); please use shading, hatching or crossing to make this figure clearer.

We respectively disagree with the review on this point. From experience, using shading or hatching generally makes things less clear as they cover the pixels of interest. We have used this approach many times before in published work (e.g. Pope et al., 2018 and Pope et al., 2020). This approach is used to show where pixels are significantly different (i.e. typically large differences) and we believe this is clear in Figure 2.

Figure S7 d: why is data not shown for Oct-Dec?

At the time of the original simulations, the meteorological reanalyses were not available and thus TOMCAT ozone was only plotted up to September 2018. However, the simulation was later extended so this available data has been included in the new Figure S7d panel.

Minor corrections

p3, l 88: ozone is repeated twice

This has been corrected.

p3, l 105: years is repeated twice

This has been corrected.

p4, l 138: Pope 2020 should be Pope et al. 2020

This has been corrected.

p4, l 141: verses should be versus

This has been corrected.

p8, l 337: change “centre panels” to “middle column”

This has been updated.

p11, l455: “Intrusion of stratospheric O3 into the mid-troposphere has a moderate influence on the observed/modelled O3 enhancements with contributions of up to 15.0-40.0%”. A contribution of 15-40% is a large contribution, not moderate.

While we agree with the reviewer that 15-40% is large in percentage terms, in absolute terms, it is moderate. Therefore, we believe the term “moderate” is suitable.

References:

Akritidis, D., Pozzer, A. and Zanis, P. 2019. On the impact of future climate change on tropopause folds and tropospheric ozone. *Atmospheric Chemistry and Physics*, **19**, 14387-14401, doi: 10.5194/acp-19-14387-2019.

Archibald, A.T., et al. 2020. Description and evaluation of the UKCA stratosphere–troposphere chemistry scheme (StratTrop vn 1.0) implemented in UKESM1. *Geoscientific Model Development*, **13**, 1223–1266, doi: 10.5194/gmd-13-1223-2020.

Churkina, G., Kuik, F., Bonn, B., et al. 2017. Effect of VOC Emissions from Vegetation on Air Quality in Berlin during a Heatwave. *Environmental Science and Technology*, 51(11), 6120-6130, doi: 10.1021/acs.est.6b06514.

Eskes, H. J. and Boersma, K.F. 2003. Averaging kernels for DOAS total column satellite retrievals. *Atmospheric Chemistry and Physics*, **3**, 1285–1291, doi: 10.5194/acp-3-1285-2003.

García-Herrera, R., Díaz, J., Trigo, R.M., et al. 2020. A Review of the European Summer Heat Wave of 2003. *Critical Reviews in Environmental Science and Technology*, **40**(4), 267-306, doi:10.1080/10643380802238137.

Graham, A.M., Pringle, K.J., Arnold, S.R., et al. 2020. Impact of weather types on UK ambient particulate matter concentrations. *Atmospheric Environment*: *X*, **5**, doi:10.1016/j.aeaoa.2019.100061.

Heal, M.R., Heaviside, C., Doherty, R.M., et al. 2013. Health burdens of surface ozone in the UK for a range of future scenarios. *Environment International*, 61, 36-44, doi:10.1016/j.envint.2013.09.010.

Honrath, R.E., Owen, R.C., Val Martin, M., et al. 2004. Regional and hemispheric impacts of anthropogenic and biomass burning emissions on summertime CO and O_3 in the North Atlantic lower free troposphere. *Journal of Geophysical Research: Atmospheres*, 109(D24), doi: 10.1029/2004JD005147.

Papanastasiou, D.K., Melas, D. and Kambezidis, H.D., 2015. Air quality and thermal comfort levels under extreme hot weather. *Atmospheric Research,* 152, 4-13, doi: 10.1016/j.atmosres.2014.06.002.

Pimlott, M.A., Pope, R.P., Kerridge, B.J., et al. 2022. Investigating the global OH radical distribution using steady-state approximations and satellite data. *Atmospheric Chemistry and Physics,* **22**, 10467-10488, doi: 10.5194/acp-22-10467-2022.

Pope RJ, et al. 2018. Widespread changes in UK air quality observed from space. *Atmospheric Science Letters*, **19:e817**, doi: 10.1002/asl.817.

Pope RJ, et al. 2020. Substantial Increases in Eastern Amazon and Cerrado Biomass Burning-Sourced Tropospheric Ozone. *Geophysical Research Letters*, **47 (3)**, e2019GL084143, doi: 10.1029/2019GL084143.

Rasilla, D., Allende, F., Martilli, A., et al. 2019. Heat Waves and Human Well-Being in Madrid (Spain). *Atmosphere*, 10(5), 288-309, doi: 10.3390/atmos10050288.

Rennenberg, H., Loreto, F., Polle, A., et al. 2006. Physiological Responses of Forest Trees to Heat and Drought. *Plant Biology*, **8**(5), 556-571, doi: 10.1055/s-2006-924084.

Richards NAD, et al. 2013. The Mediterranean summertime ozone maximum: global emission sensitivities and radiative impacts. *Atmospheric Chemistry and Physics*, **13**, 2231-2345, doi: 10.5194/acp-13-2331-2013.

Rodgers, C.D. 2000. Inverse methods for atmospheric sounding: Theory and practice. New Jersey, USA: World Science.

Roelofs, G.L., Kentarchos, A.S., Trickl, T., et al. 2003. Intercomparison of tropospheric ozone models: Ozone transport in a complex tropopause folding event. *Journal of Geophysical Research*, **108** (D12), 8529, doi: 10.5194/acp-19-14387-2019.

Schultz, M.G., Schroder, S., Lyapina, O., et al. 2017. Tropospheric Ozone Assessment Report: Database and metrics data of global surface ozone observations. *Elementa – Science of the Anthropocene*, **5**(58), doi: 10.1525/elementa.244.

Stirling, E.L., Pope, R.J., Graham, A.M., et al. 2020. Quantifying the transboundary contribution of nitrogen oxides to UK air quality. *Atmospheric Science Letters*, **21**(2), doi: 10.1002/asl.955.

Stockwell, D., Giannakopoulos, C., Plantevin, P.-H., et al. 1999. Modelling NO_x from lightning and its impacton global chemical fields. Atmospheric Environment, **33**, 4477–4493, doi: 10.1016/S1352-2310(99)00190-9.

Yan, Y., Pozzer, A., Ojha, N., et al. 2018. Analysis of European ozone trends in the period 1995—2014. *Atmospheric Chemistry and Physics*, **18**(8), 5589—5605, doi: 10.5194/acp-18-5589-2018.