

Responses to reviewers

In this document we have presented a point-by-point response to the reviewers' comments. The original comments by the reviewer are shown in black while our responses are shown in blue.

Reviewer 1:

This manuscript attributes trends and seasonal cycles in ozone metrics (PSO and MDA8) across Europe and North America to NO_x and VOC sources. Some key findings include that declining local emissions of NO_x contribute to decreased O₃ production in summertime but increasing import of foreign precursors leads to increasing winter and springtime O₃; that natural VOC and local anthropogenic NO_x are often linked in their formation of O₃; and that foreign NO_x is now of similar magnitude to local NO_x in terms of contribution to O₃ formation in several locations.

These findings are achieved using a model in which the precursor molecules are tagged by source and location. This is a considerable amount of work, which has yielded novel results. The conclusions are of interest to the community and the work fits well into the scope of the journal and special issue.

We thank the reviewer for their positive comment - it was indeed a lot of work but we are glad that it has produced some novel results that are of interest to the community.

Although long, the text is clearly written and the conclusions are well supported by the data. I would recommend this manuscript for publication with only a few amendments/suggestions.

General comments

The methodology is very clear on several aspects such as the emissions inventories but no detail on the chemical model itself – how many species and reactions, how are VOC and oxidation products treated / lumped? The text refers to MOZART but it's never explicitly introduced that you are using this scheme.

We have now added a few sentences on the chemical mechanism and treatment of VOCs and also pointed the reader to the relevant literature for further details on the MOZART chemical mechanism: L183-190:

“The gas-phase chemical mechanism employed in this study is based on the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4) (Emmons et al., 2010) which includes detailed O_x-NO_x-HO_x-CO-CH₄ chemistry, along with the oxidation schemes for a range of non-methane volatile organic compounds (NMVOCs). Specifically, MOZART-4 treats 85 gas-phase species involved in 39 photolytic and 157 gas-phase reactions. NMVOCs are represented using a lumped species approach, where, for example, alkanes larger than ethane are lumped as a single species (e.g., BIGALK for C₄+ alkanes), and alkenes larger than ethene are lumped (e.g., BIGENE), with specific treatments for aromatics, isoprene, and terpenes. The oxidation products of these lumped

and explicit VOCs are also tracked. Further details on the MOZART-4 chemical mechanism, including the full list of species and reactions, can be found in Emmons et al. (2010)”

The structure of the manuscript is ok as is but perhaps could be improved. Some of the results sections begin with several large paragraphs of introduction before any results are given. Some of this text may be better placed in, or is perhaps a repeat of, the methods sections. Similarly, descriptions of figure 2 would work in the results rather than the methods.

We considered including the emissions trends in the results section but felt that it was more appropriate to cover all descriptive aspects of model input, including regional and sectoral emission trends, in the model description section. Since emissions are not an output from the model runs, we do not treat them as part of results. However, we do refer back to these emission trends in relation to modelled ozone in the results section when necessary. We have now included new spatial maps of anthropogenic NO_x emissions for North America and Europe in the result section (Figures 5 and 11).

The individual regions are described independently but never compared. Are there any key differences between regions that could be highlighted in the conclusion? If not, do all the regions need to be described in such detail?

Thank you for your comment. This point has also been raised by reviewer 2 and we have substantially rewritten sections 3.2 and 3.3 with more emphasis on intercomparison of same source contributions across all receptor regions for a given continent. We have also made the discussion less qualitative and redundant.

The difference in emissions reductions between western and eastern Europe is highlighted as a reason why PSO trends may differ regionally but this hypothesis is never confirmed. There don't seem to be large differences in PSO changes over time between the regions. Is that correct? If so, can you suggest why? A follow-up on this introductory point is needed somewhere.

We have now included new spatial maps of the peak season anthropogenic NO_x emissions, PSO, and local NO_x contribution to PSO for the initial (2000) and final (2018) year for both North America and Europe. See Figures 5 and 11. Figure 11a and 11d clearly show the relatively small change in anthropogenic NO_x emissions in Central & Eastern Europe over the 19-year period as compared to Western Europe. Section 3.3 has been rewritten now and section 3.3.1 discusses this issue more specifically in lines L1109-1118

“To understand the geographical backdrop of PSO changes, Figure 11 presents a spatial map of local anthropogenic NO_x emissions (panels a, d), total PSO (panels b, e), and the modeled contribution of local anthropogenic NO_x to PSO (panels c, f) for the initial (2000) and final year (2018). In 2000 (Figure 11a), prominent NO_x emission hotspots were evident (e.g., Benelux, Germany, Po Valley), parts of the UK, and major urban agglomerations across the continent. By 2018 (Figure 11d), substantial emission reductions occurred, particularly in Western and Central Europe. However, this decline is not obviously reflected in the spatial patterns of total PSO (Figure 11b, e), which generally decreased in the southern regions but not in northern regions, especially over areas with the largest emission cuts, as also seen in bias-corrected PSO maps by Becker et al. (2023). The direct contribution of local anthropogenic NO_x to PSO (Figure 11c, f) mirrors these emission reductions

more closely, with clear reductions from 2000 to 2018. This suggests the role of other contributions in offsetting the expected decline in PSO, especially in northern European regions.”

The conclusion could be developed somewhat, for example by comparing the different regions and Europe vs the US (the US emissions reductions seem more successful at reducing O₃?) as stated above. There could also be greater discussion on some of the uncertainties – are emissions inventories the greatest source of uncertainty in your study? I would finally also be interested in your thoughts on how this novel tagging approach could be used in future studies – could it play a role in bias correction and identification?

We have now substantially updated the Conclusion, Limitations, and Future Outlook section. It now covers some comparative aspects of North American and European changes in precursor emissions and observed ozone, a couple of statements on the potential sources of biases (model resolution, emission inventory, treatment of ship plumes, deposition schemes), and future directions for the use of this tagging approach.

Data should be made publicly available according to Copernicus guidelines

We will make the data produced in this study publicly available before the final publication of the manuscript.

Units and trend analysis should follow the TOAR special issue guidelines. Just check if any of those are relevant to this work.

We have updated our trend analysis to include p-values and 95% confidence intervals based on the TOAR statistical guidelines.

Specific comments:

L26: A concluding sentence could be useful here.

We have now added a concluding sentence “Our results highlight the evolving drivers of surface ozone and emphasize the need for coordinated global strategies that consider both regional emission trends and long-range pollutant transport.”

L45: In case you would like the more recent publication: Cheesman, Alexander W., et al. "Reduced productivity and carbon drawdown of tropical forests from ground-level ozone exposure." *Nature Geoscience* (2024): 1-5.

We have now included this citation.

L56: It is worth being clear that the O₃ precursors can also be transported, since this is the key part of the study.

Thank you for this suggestion. We have added a new sentence in L70-72: “Moreover, some ozone precursors (e.g., CO and less reactive NMVOCs) also possess atmospheric lifetimes sufficient for

intercontinental transport, subsequently contributing to ozone formation in downwind regions far from their original emission sources.”

L210-214: This is slightly hard to follow and explained much better in Table1 and the associated table caption. I would suggest you direct readers to this table earlier. I would also suggest on L213 to replace ‘many zones’ with the specific number of zones.

We have now directed the reader to Table 1 earlier, and have modified L252-253 to “*The ocean is also divided into multiple zones and tagged separately (see Figure S12).*” Figure S12 is a new figure which shows these oceanic zones.

L230: if you plan to refer to anthropogenic VOCs as AVOCs this could be introduced earlier and then used consistently throughout (such as in the figure labels)

We have now replaced subsequent instances of “*anthropogenic NMVOCs*” with “AVOC”

L226-250: This reads as results rather than methods to me. Certain aspects such as AVOCs showing an increasing trend from 2013 and global lightning NO_x showing a decreasing trend are interpretations of the figure. Consider moving this to the results or giving its own subheading in the methods section.

We decided not to move this description of emissions since it is not a model output. However, we have revisited emissions distribution vis-a-vis PSO distribution in the results section (3.2.1 and 3.3.1) through the addition of new figures (Figures 5 and 11).

Section 2.2 could be condensed. The definition of MDA8 is essential but the wallclock time in your specific setup could be left out.

We decided to retain this piece of technical information because we feel it allows other researchers to compare the computational advantage of using a tagging-based system (such as TOAST 1.0) against, say, a set of multiple perturbation simulations which may be used to derive similar source contributions. We believe documenting such computational requirements is scientifically valuable and particularly helpful for new researchers entering the field who often struggle to find such information in published literature.

In figure 3, is it possible to just show the source regions being considered? It is a bit unclear which 5 regions of N America and Canada are being selected at this stage.

Thank you for this suggestion. We have now updated Figure 3 such that only the receptor regions discussed in this study are shown. The map has been cropped out to enlarge the discussed regions and the colour scheme has also been updated to emphasize the different sub-regions within a continent.

L305: Inhomogeneous measurements combined with changing numbers of stations over time can lead to errors in mean and trend identification. An acknowledgement of the uncertainty in

the observational trends would be helpful here. Christoph Frei has done a lot of work on this for temperature and precipitation fields.

We have now added the following sentence with an appropriate reference L360-361: *“We note that sparse spatiotemporal sampling can introduce uncertainty in identifying true long-term trends of ozone and refer the reader to a technical note by Chang et al. (2024) for more details on this issue.”*

L382: I agree the simulation of MDA8 against rural stations is very nice. I am not sure what the satisfactory performance across different world regions refers to.

We have now made this evaluation more holistic by also adding annual averaged correlations (to remove the influence of O3 seasonal cycle in driving up the r-values) and mean absolute bias (MAB) to reduce the compensatory effects of high and low biases. These numbers compare quite favorably to other studies (also added now).

L387: What is a 19 year month centered average MDA8 O3 broken down into 5 years?

Thanks for pointing this out. We have now made this sentence clearer: *“We also present the first five year (2000-2004) and last five year (2014-2018) month-centered average MDA8 O3 seasonal cycle and explain the shifts in terms of tagged contributions for all receptor regions during these periods.”*

L406: A reminder of the PSO guideline could be useful here.

We have now included the WHO PSO guideline values (31 ppb) here.

L444-446: Is it possible for the PSO to be ‘single handedly’ linked with local AVOC and then linked to declining NOx emissions in the following sentence?

We have removed *“almost single handedly”* from this sentence.

L525: I would pedantically argue that all ozone is equally destroyed by water, but long-range transport ozone is more likely to encounter it.

This section is entirely rewritten but we have made this point clearer in the new text in lines L881-885: *“The summertime dip in foreign NOx contribution (also seen in other sub-regions) is likely due to shorter lifetime of ozone at higher temperatures, which is associated with increased water vapor content in the atmosphere (Stevenson et al., 2006). Water vapor promotes ozone loss via photochemical pathways involving HO_x radicals, and transported ozone is more likely to be destroyed under moist conditions (Real et al., 2007). Consequently, the efficiency of long-range ozone transport decreases in summer”*

L558: perhaps refer back to the figures here

This section is rewritten.

L738: Can you say why local NOx and BVOCs are so interlinked? This is an interesting finding.

We have added more text to emphasize this finding: L1410-1421

“Due to the nature of our ozone tagging system, we perform two separate source attributions, one for NOX emissions, and another for VOC emissions. When attributing ozone to VOC emissions, we note the strong contribution of BVOC emissions to the summertime peak ozone, which is clearly linked with the strong contribution of local anthropogenic NOX emissions to summertime ozone. The co-variability of these two sources is also apparent in the PSO time series for all regions and emphasizes the interaction of anthropogenic NOx with BVOC in rural and background regions. This is an emerging finding made possible due to our dual-tagging approach; a relatively recent regional modelling study (Lupascu et al., 2022) focusing on two high ozone episodes in Germany that also utilized the TOAST1.0 system also noted the interaction of local anthropogenic NOx and BVOC in driving ozone peaks.and emphasizes the interaction of anthropogenic NOx with BVOC in rural and background regions. This finding highlights that, at least for rural and background regions, the interaction of anthropogenic NOx with BVOC exceeds its interaction with AVOC which might be contained within the urban centres. It is noteworthy that BVOC emissions also either match or exceed AVOC emissions in North America and Europe during the peak season.”

Technical points:

L22: ‘productivity’ could be replaced with ‘O3 production’ or similar for clarity.

We have now made this amendment.

Figure 7 b/d is mislabelled in the figure caption. Same for other figures of this layout.

Thanks for pointing this out. We have now corrected this in the captions.

Figures 2 and 3 are sometimes referred to incorrectly in the text (e.g. Fig 2 written instead of Fig. 3 and vice versa)

We have now corrected this issue.

=====

Reviewer 2

This paper presents a global modelling study with an innovative dual tagging analysis; the paper's focus is on surface ozone over North America and Europe, particularly with regard to the causes of long-term changes and the seasonal cycle and its changes. In my view, the work appears to employ a state-of-the art modeling system and addresses an interesting scientific issue.

We thank the reviewer for their positive comment on the scientific relevance of our research question and the innovative nature of our methodology.

However, ACP aims to publish studies with important implications for our understanding of the state and behavior of the atmosphere; I believe that the paper at present fails to advance this aim. Thus, I recommend that this submission be rejected and the paper returned to the authors with encouragement to resubmit if the authors can address 3 critical issues:

Significant questions remain both regarding the evaluation of the modeling system and the possible corruption of observational data upon which the model evaluation and analysis are based.

A more quantitative approach to analysis would improve the paper, both with regard to discussion of present results and discussion of comparisons and contrasts with previously published analyses.

A clear summary is required of what new understanding of the atmosphere has emerged from this study, including the added value of the dual tagging approach, and when and for what issues that approach is required in such modeling studies.

We thank the reviewer for seriously engaging with our work and pointing out these caveats which we have now addressed in the comprehensively revised manuscript.

More detailed discussions of these issues follow. In addition, several related and unrelated, major and minor issues are discussed that may be of use to the authors for their revision.

Critical issues:

Section 3.1 is devoted to evaluation of the CAM-Chem model used in the chemical-transport simulations. In my view a much more robust evaluation is required as outlined in the following paragraphs; I believe that, in general, such evaluation is necessary before model simulations can be relied upon to provide robust results.

One aim of the authors is to explain the seasonal cycles and their changes observed in surface ozone; thus the model evaluation should move beyond an overall statistical comparison of monthly mean concentrations between model and observations; it is necessary to specifically and quantitatively evaluate how well the model reproduces the phenomena of interest, in this case the seasonal cycles and their changes.

We have now adopted a more quantitative approach to model evaluation with a particular focus on the reproduction of seasonal cycles (of individual years as well as multi-year averages) over various receptor regions as captured by fourier parameters (amplitude and phase angles of constituent harmonics).

The three paragraphs on lines 339–372 discuss aspects of such an evaluation, but only qualitatively; this evaluation should be placed on a quantitative basis. Of concern are the findings in previously published intercomparisons of model simulations with observed seasonal cycles of ozone. Parrish et al. (2016) discuss such intercomparisons for marine boundary layer (MBL) sites and in the overlying free troposphere above one site. Three chemistry climate models, including a version of the CAM-Chem model used in the present study, approximately reproduced many features of the measured seasonal cycles within the MBL, with some notable quantitative disagreements, but gave divergent results that do not agree with measurements above the MBL. Bowman et al. (2022) discuss a similar intercomparison that considers both the seasonal cycles and their systematic changes at northern midlatitude baseline locations. The available observational data were compared with simulations by 6 Coupled Model Intercomparison Project Phase 6 (CMIP6) Earth system models, again including a version of the CAM-Chem model. Substantial differences were found between the different model simulations and between the simulations and the observations. To my knowledge, the model disagreements revealed in these intercomparisons have not been addressed in later model development. A quantitative evaluation such as presented in these two papers should be included in the present manuscript.

We have now adopted the approach of Parrish et al. 2016 and Bowman et al. 2022 and broken down the modelled and observed seasonal cycles over all the receptor regions considered in our study into two harmonics using a fourier transform, thus yielding values such as the detrended y-intercept and the amplitudes and phases of the two harmonics which allow for a more quantitative comparison of various features of the observed and modelled seasonal cycles. We applied this approach to the seasonal cycles for each individual year (2000 to 2018) as well as time-averaged seasonal cycles over the first five (2000–2004) and last five (2014–2018) years and also the full 19-year (2000–2018) averaged seasonal cycles. These parameters are listed in tables S1–S9 for different receptor regions. We have discussed the model performance in terms of these parameters both in the model evaluation section (section 3.1) as well as in the analysis sections (3.1, 3.2.2, 3.2.3, 3.3.2, 3.3.3)

The authors primarily rely on an overall statistical analysis of the agreement between model simulations and observations of monthly mean ozone in specific regions. Their final conclusion (line 374) is: “Overall, we obtain very good model–observations agreement, with low biases and high correlations,” Three issues must be addressed regarding this conclusion. First, this evaluation is limited to comparison between model results and observations of time series of MDA8 ozone that are highly averaged, both temporally (monthly) and spatially (first over model grid cells and then over receptor regions); it should be discussed if this averaging of model result is appropriate in the context of the model results that the authors employ in the following discussion of results.

Since all our subsequent analyses and conclusions depend on Peak Season Ozone values averaged over the defined receptor regions, which in turn depend on monthly mean MDA8 O₃ values and not on any high frequency extremes, evaluation of monthly mean MDA8 O₃ over the different receptor regions is adequate and appropriate in this context. We have now added a line in our evaluation section highlighting this point.

L493–496: *“We note that our model evaluation is based on model results and observations of time series of MDA8 O₃ that are averaged, both temporally (monthly) and spatially (first over model grid cells and then over receptor regions) but such an evaluation is valid because all our further analyses and conclusions depend on the same spatial and temporal scales.”*

Second, for the 10 receptor regions discussed in the paper (shown in Figure 4) and 5 additional receptor regions (shown in Figure S1) the average of the mean biases is indeed small (1.75 ppb), but more context is required for these mean biases. The regional mean biases range from -9.57 to 8.84 ppb, giving an overall regional mean absolute bias twice as large (3.50 ppb) as the overall mean bias. These statistics provide some evaluation of model performance in simulating average ozone concentrations, but the paper aims to quantify ozone “trends and changing seasonal cycles”. It is clear from examination of Figures 4 and S1 that there are significantly larger deviations between the observations and model results for the individual monthly means than are reflected in the regional mean biases. The authors should give a more detailed view of the time series comparison including discussion of an additional statistic (I suggest mean absolute difference), which would more relevantly quantify the differences of monthly means between model results and the observations – it is these differences that are of most importance with regard to comparison of seasonal cycles and their changes.

We have now also calculated the Mean Absolute Bias (MAB) which are printed over each panel in Figures 4 and S1. We have also discussed the MAB in the text in section 3.1.

Second, the high correlations the authors cite (generally $r > 0.9$) are a) expected and b) not universal (3 of 15 r values are 0.62 or smaller); large r values are expected simply because the variability in both model results and observations is dominated by large seasonal cycles that are approximately in phase; if annual averages were compared, the correlations would be much lower. Examination of the figures strongly suggests that the smaller r values very likely indicate spurious observational data; the origin and influence of these spurious data must be assessed as discussed further below.

We have now also included correlation coefficients between annual averaged observed versus modelled MDA8 O₃ (shown in brackets in figures 4 and S1). These are indeed smaller than

r-values between observed vs modelled monthly mean MDA8 O₃ because they essentially denote interannual variability between modelled and observed ozone which to a large extent depends on year-to-year meteorological changes. These are also the same as correlation coefficients between the y_0 which are further presented in Figures S3 and S4. Section 3.1 has been extensively rewritten with a focus on the model's skill in reproducing the observed Fourier parameters (y_0 , A_1 , A_2 , ϕ_1 and ϕ_2).

Third, from a skeptical viewpoint, we must be mindful of a subtle issue associated with model-observation comparisons such as the authors present in Section 3.1 (and that are also utilized in many such comparisons in the published literature). Chemical transport models do not treat all relevant processes from first principles of physics and chemistry; rather there are a great many parameterizations embedded within the computer code of the models. Over past decades those parameterizations were developed and tuned so that the models reproduce mean observed ozone concentrations as closely as possible. Consequently, attempting to evaluate the performance of models simply from comparison of means of observations with model results involves a degree of circular reasoning – the models were developed to agree with observations, so such agreement cannot be taken as independent confirmation that models perform properly for the correct reasons.

We acknowledge the general concern that parameterizations within models are often tuned over time to improve agreement with observed mean concentrations, which could introduce a degree of circularity in model evaluation. However, we believe several aspects of our specific evaluation mitigate this concern and demonstrate the robustness of our findings for CAM4-Chem using TOAR-II gridded observations: First, our evaluation focuses on the Maximum Daily 8-hour Average (MDA8) O₃, which is a health-relevant peak metric. Reproducing MDA8 O₃ accurately requires the model to capture not only mean ozone levels but also the diurnal cycle of ozone, which is influenced by complex interactions between precursor emissions, photochemistry, boundary layer dynamics, and short-term meteorological variability. This is a more stringent test than matching simpler mean ozone concentrations, which are often the primary target for broader model tuning efforts. It is less likely that model parameterizations were specifically tuned to universally optimize MDA8 O₃ across diverse global regions.

Second, we use the recently released TOAR-II gridded surface ozone product for evaluation. While CAM4-Chem, like other global models, has undergone development and validation against various observational datasets over its lifetime, it is unlikely that its parameterizations were specifically and extensively tuned to reproduce the regional monthly mean MDA8 O₃ values derived from this particular TOAR-II gridded dataset. The TOAR-II dataset is based on a comprehensive collection of global observations, processed and gridded, offering a relatively independent benchmark for models developed prior to or in parallel with its creation. The version of CAM4-Chem used in this study has a development history that would have drawn on a broader range of observational inputs, likely not focused explicitly on optimizing for the TOAR-II MDA8 O₃ metric.

Third, our analysis demonstrates high correlations and low biases for regionally averaged monthly mean MDA8 O₃ values across all evaluated receptor regions. Achieving such consistent performance across multiple diverse regions, each with unique emission profiles, chemical regimes, and meteorological conditions, suggests that the model captures key underlying processes governing ozone production, transport, and loss with reasonable fidelity. This goes

beyond simply matching a global mean and indicates skill in representing regional-scale phenomena and their seasonal variability.

Finally, the treatment of ship emissions and possibly the MBL structure constitute one apparent shortcoming in the modeling methodology that must be addressed; I believe this is a wide-spread shortcoming of chemical transport models, and has significant impact on this study's results. The model treatment of ship emissions is not discussed in detail in Section 2, but the authors do quantify ozone produced from this emission source. Unless the model includes some sort of plume-in-grid treatment for ship plumes, it must be expected that the influence of ships is significantly exaggerated (e.g., Kasibhatla et al., 2000). I am aware of only a single observational study that provides a detailed characterization of ship plume evolution (Chen et al., 2005). This study shows that the photochemical perturbation of the marine boundary layer (MBL) due to a ship plume is largely limited to the first few hours of plume evolution, while the plume is still narrowly confined (FWHM ~ 3 km). Thus, an effective model resolution < 1 km is required to accurately treat ship emissions. However, I suspect that the model utilized in this study (as well as many other studies) immediately distributes the ship emissions throughout the 1.9°×2.5° model cell; this is expected to lead to a large overestimate of ozone production from ship emissions.

We thank the reviewer for this insightful comment regarding the treatment of ship emissions and their impact on ozone production. The reviewer is correct that our CAM4-Chem configuration, consistent with many global chemical transport models, does not employ a plume-in-grid parameterization for ship emissions. Instead, these emissions are instantaneously diluted within the 1.9° x 2.5° model grid cell. We acknowledge that this approach can lead to an overestimation of ozone production efficiency from ship NO_x compared to the more localized, high-NO_x conditions within a concentrated young plume where titration effects and NO_x self-reactions can be more dominant. The immediate dilution into a larger volume with potentially lower background NO_x can artificially enhance the per-molecule ozone production from the emitted ship NO_x. Therefore, while our study quantifies the contribution of ship emissions as resolved by the model, we recognize this as a potential uncertainty, and the true ship NO_x contribution might be somewhat lower than simulated. Addressing this accurately would indeed require sub-grid scale treatments, which are computationally intensive and not yet standard in all global modeling frameworks. We have added new text to our discussion of limitations to reflect this point:

L472-485: “We also evaluate the model in the context of potential overestimation of ozone production from ship plumes. This is because in our modelling setup, ship NO_x emissions are instantaneously diluted within the 1.9°×2.5° model grid cell which can lead to an overestimation of ozone production efficiency from ship NO_x. In the real world, the more localized, high-NO_x conditions within a concentrated young plume, the titration effects and NO_x self-reactions can be more dominant and the true ship NO_x contribution might be somewhat lower than simulated (Kasibhatla et al., 2000; Chen et al., 2005; Huszar et al., 2010). Such overestimated ship NO_x contribution to ozone shows up, for example, in terms of a lower simulated vertical gradient than the observed vertical profile of ozone especially at remote coastal locations. To assess this, we plot observed and model simulated ozone vertical profiles at Trinidad Head, off the coast of California, for the month of July (a representative month for peak season) for all 19 years (see Figure S5). The

monthly mean modelled vertical O₃ profile over Trinidad Head generally falls within the envelope of daily observational profiles within the MBL (say, below 850 hPa). Although, for multiple years, the vertical drop in modelled O₃ concentration towards the surface is less sharp than that seen in observations, thereby suggesting a potential overproduction of O₃ near the ocean surface in the model due to instantaneous distribution of ship NO_x emissions in the model gridcell. This particular feature of our modelling system can partly explain the positive bias in simulated ozone."

One zeroth-order check that the authors should perform is a comparison of their total marine ozone production from ship plumes with their total ship emissions NO_x during the photochemical active season of the year; approximate agreement with a 4.25:1 mole/mole ratio is expected (i.e., production of 10 O₃ molecules per NO_x emitted during the day, based on the Chen et al., 2005 study), and loss of 1.5 O₃ molecules per NO_x emitted during the night, assuming that NO_x is lost as N₂O₅ at night). Duncan et al. (2008) further discuss this issue. This comparison should be limited to ozone production from ship emissions within the MBL at northern mid-latitudes.

We tried performing this sanity check with the output data available to us. However, there were a number of obstacles to ascertaining the suggested 4.25 : 1 O₃-from-ship : ship-NO_x molar ratio as suggested by the reviewer. First, we only have monthly mean emissions from each source in the input and therefore cannot explicitly distinguish daytime and nighttime NO_x emitted from the ships. Similarly, due to storage constraints, we do not output 3D ozone from different tags, including ships, at the hourly frequency. We only output surface ozone from each tag at the hourly level and 3D ozone from all tags at the monthly mean level. Therefore, we could only perform this MBL-wide analysis at the monthly mean scale. Using July 2018 as a representative month for the peak season, we calculated the total NO emitted from the shipping sector over northern hemisphere midlatitudes (30°N–60°N) to be **1.138e+10 moles**. We then utilized the monthly mean 3D ozone concentration field from the ship NO_x tag (raw output being in mol/mol-of-dry-air) to calculate moles of O₃ attributed to ship NO_x over July 2018. We found the total global O₃ attributed to global ship NO_x to be **5.308e+11 moles**, O₃ within the MBL attributed to global ship NO_x to be **8.703e+10 moles**, and O₃ within the northern hemisphere midlatitude (NHML) band MBL attributed to global ship NO_x to be **8.019e+10 moles**. The ratio of NHML MBL ozone produced from ship NO_x and the ship NO_x emissions over NHML turns out to be **8.019e+10/1.138e+10 = 7.04** which is higher than the expected ratio of 4.25 as suggested by the reviewer. However, we can immediately see the issue with this approach: here we have a large number of moles of NO in the denominator which were arrived at by multiplying the per second emission rate with the number of seconds in a month (a large number), while the numerator was governed by the monthly average O₃ concentration. If we perform the same calculation on a daily scale, the numerator won't change much (a daily O₃ concentration maybe quite similar to a monthly mean O₃ concentration) but the denominator (i.e. moles of NO emitted by ships in a day) would be substantially smaller, thereby making the ratio much larger. Therefore, a reliable sanity check could only be performed if we had tagged ozone production rates rather than tagged ozone concentrations. Unfortunately, our current tagging system does not provide ozone production rates from different tags so it is not feasible to estimate the molar ratio as suggested by the reviewer.

A second zeroth-order check that the authors should perform is a comparison of their simulated vertical profile of ozone within and directly above the MBL with observations. Ozone sonde data from Trinidad Head (e.g., Fig. 15, Oltmans et al., 2008, Fig. 12, Parrish et al. 2016 or Fig. 1, Parrish et al., 2022) are available; these represent the marine environment, so they should be compared with model simulations from a grid cell offshore of northern California. Note that Fig. 12 of Parrish et al (2016) compares the measured vertical gradient of ozone with simulations from 3 global models, one of which is a version of CAM-Chem model that was used in the present paper. Importantly, the observed strong near-surface vertical gradient of ozone over the ocean clearly indicates that the MBL is a region of strong domination by ozone loss; a spuriously large ozone source within the MBL, such as overestimate of the ship emission source, would be expected to disrupt the relatively strong vertical gradient through that level. Notably, none of the 3 models reproduced the observed ozone gradient within the MBL.

We obtained the ozone sonde data over Trinidad Head for the 2000-2018 period and were able to perform this sanity check. As mentioned earlier, we only store 3D ozone data at the monthly mean level from the model output. The sonde data were available at a daily level for a varying number of days per month over the 19 year period. We plotted the daily observed ozone profiles over Trinidad Head for the peak season representative month of July for the 2000-2018 period along with the corresponding model-derived monthly averaged ozone profiles (see Figure S5). We found that the monthly mean modelled vertical O₃ profile over Trinidad Head was generally within the envelope of daily observational profiles within the MBL (say, below 850 hPa). Although, for multiple years, the vertical drop in O₃ concentration towards the surface was less sharp than that seen in observations, thereby suggesting an overproduction of O₃ near the ocean surface in the model due to instantaneous distribution of ship NO_x emissions in the model gridcell. We have now discussed this in lines 472-485 and have qualified our conclusions relating to the contribution of the shipping sector to surface ozone (lines 1393-1394).

There are also indications of possible corruption of observational data relied upon in this paper. As noted earlier, Figures 4 and S1 include data that disagree strongly with model simulations and are simply physically unreasonable (e.g., too large in Fig. S1d and too small in Fig. 4i). Further, the authors discuss an anomalous drop in 2012 in the number of rural TOAR stations. Figure S8 shows that this drop was only in the US, not Europe, although southern Europe shows a peak in 2012 and western Europe shows a drop in 2017, both of which are unexplained. Northeastern US also shows a rather large increase in site number after 2015. The US ozone data are available from the US EPA data archive; that archive does not include the TOAR site classification, but the total number of US sites reporting ozone observations increased from 1241 in 2000 through 2010 and remained relatively constant at about 1450 for the three 2011-2013 years, followed by an accelerating decrease to 1231 by 2024. There is no indication of a drop in the number of stations in the US in 2012. Since the US monitoring network remains relatively constant from year-to-year with only small numbers of stations coming on line or closing, and with no systematic movement of sites to or from rural areas, it is clear that the data that the authors extracted from the secondary TOAR archive does not accurately reflect the parent EPA archive from which the TOAR data were obtained. These two features of the observational data make it imperative that the source of these problems be tracked down, and the observational analysis included in this paper be thoroughly evaluated and revised as necessary before included in a submitted manuscript.

We thank the reviewer for pointing this out. We further investigated the anomalous drop in the no. of stations in the US for 2012. It turned out that the data retrieval scripts for the TOAR-II gridding database had returned errors for many US stations for 2012 which were ignored by the automated retrieval system ultimately leading to many missing US stations for 2012. We now use an updated database which includes all valid stations for 2012. For SE Europe, the data for initial years comes from a single station, Aliartos, in Greece, which is a coastal location and may not be representative of the larger gridcell sampled. For southern Africa too, the data comes from a single station and there is lack of clarity regarding units (ppb or ug/m³) for the year 2014 from the original data providers. We have decided to exclude the observed data for 2014 for southern Africa. The slight increase seen in the number of stations in NE US is due to the inclusion of new non-EPA stations from the OpenAQ portal (<https://openaq.org>). We have also updated the station weighting mechanism in the TOAR-II gridding system. For certain stations, there were multiple (two or three) sensors placed at the same site which provided independent ozone observations that led to over-weighting of those stations. Now, these time-series are first averaged to produce a single time-series per station. Thirteen new observational time-series were included in the updated dataset which were earlier ignored due to a special character in their filename. We have created a new time series plot of the number of valid stations per receptor region (Figure S11) using the updated TOAR-II dataset which now turns out to be free from any dramatic peaks and drops.

Since the observed data has changed significantly, we have also updated the model output used for further analyses by performing a fresh co-sampling using the new TOAR-II dataset. Regional averaged MDA8 O₃ and PSO values were recalculated and re-plotted for all receptor regions. All figures in the new manuscript now use these new model-extracted values co-sampled against the updated TOAR-II dataset.

Sections 3.2 and 3.3, which comprise nearly half of the paper, present and discuss the results of the study. However, the discussion is largely a qualitative catalog of features apparent in the observations or model results. That discussion should be extensively revised to replace that qualitative catalog with more systematic and quantitative analysis approaches of the model results, comparisons between model results and observations, and between results in different receptor regions. The separation of long-term changes from the seasonal cycle (i.e., detrending the data before analyzing the seasonal cycle) is often very important, but the trends in ozone over the 2000-2018 period are so weak that this is not essential in this study. A Fourier series or spectral analysis, such as used by Parrish et al. (2016), Bowdalo et al. (2016) and Bowman et al. (2022) is recommended.

We thank the reviewer for their valuable comment. We have now entirely rewritten sections 3.2 and 3.3 with more emphasis on intercomparisons of contributions from the same source across different receptor regions. We have also discussed the observed and modelled seasonal cycles in terms of their fourier parameters.

Sections 3.2.1, 3.2.2 and 3.3.1 are primarily successive, isolated discussions of the ozone contributions in successive receptor regions, and much of this discussion is repetitive between regions. The authors provide little context for this discussion, so the reader is faced with unconnected qualitative descriptions and numbers; an improved organizing context is needed. I suggest that the authors successively discuss each of the ozone contributions over all receptor

regions in these sections, much as done in Section 3.3.2. For example, it would be informative to compare and contrast the anthropogenic NO_x contributions to ozone in the receptor regions.

We have now rewritten these sections with more emphasis on contrasting contributions from the same sector across different regions.

As expected, Figure 5 shows that contribution decreases along western North America from the Southwestern US (with many large urban areas) to the Northwestern US (with few large urban areas); in this regard, it would be useful to include the western Canada receptor region, which is not similarly discussed in the paper.

We have not included discussion for Eastern Canada due to the unavailability of TOAR-II data from rural stations in this region which prevents model evaluation and co-sampling for this region.

It would also be of interest to quantitatively examine the correlation between that ozone contribution and the total local anthropogenic NO_x emissions over all receptor regions.

We agree with the reviewer's valuable suggestion. We have now performed new analyses and created maps for North America and Europe showing the spatial distribution of the correlation coefficient r between total PSO and the local NO_x contribution to PSO over these two continents versus the local NO_x emissions. (see figures 7 and 13). We have now discussed this new analysis in sections 3.2.1 and 3.3.1 respectively for North America and Europe.

Insightful comparisons and contrasts with previously published analyses are essential. The quantitative results derived by the authors should be compared and contrasted with published results obtained by similar or differing analytical approaches to those quantifications; this further discussion must be based upon an in-depth literature review of published analyses of ozone trends and seasonal cycles and their changes over North America and Europe. Two specific examples of potential literature comparisons are summarized below; these examples should be considered as illustrative, but not as a comprehensive list of needed discussion topics.

We have now cited some previous studies (eg., Becker et al., 2023; Simon et al., 2024) to put our PSO results in context. We have also cited studies such as Bowdalo et al., 2016; Parrish et al., 2016 and Bowman et al., 2022 when presenting the quantitative results from the Fourier analyses.

Of great interest would be to fit the temporal evolution of the local anthropogenic contribution in each receptor region to an exponential function, rather than the linear analysis the authors employ. Parrish et al. (2025) and papers cited therein have shown that local anthropogenic enhancements of surface ozone in North American regions have decreased exponentially with a time constant of 21.8 ± 0.8 years. From similar analyses, Derwent and Parrish (2022) report exponential time constants for the local anthropogenic contribution of 18 ± 4 years over the United Kingdom and 37 ± 11 years over continental Europe. Comparison of the present model results to those observationally derived results would be quite useful.

We decided not to fit exponential functions to the local anthropogenic contribution to PSO because we concluded that while the exponential fit as used in Parrish et al. (2022) and Parrish et

al. (2025) was better suited for capturing the the longer, fuller, time series of ozone which begins in the 1970s with an initial increase, peaks and then declines, a linear fit sufficiently captures the broadly secular decline in ozone over the shorter 19-year period considered in our study. However, we have now contrasted our tagged local NO_x and foreign contributions to ozone to similar contributions derived through statistical approaches in published literature (see L626-645).

“Our model-based findings of declining local anthropogenic contributions to PSO in North America differ quantitatively with recent observation-based studies such as Parrish et al. (2025), which also document a significant waning of local influence using different metrics and inferential techniques. For example, Parrish et al. (2025) estimate a local anthropogenic enhancement to Ozone Design Values (ODVs) in the SW US of typically <6 ppb in recent years. Our direct tagging method quantifies a larger local anthropogenic NO_x contribution to average PSO in this region (~16 ppb in 2014-2018, Figure 6h). This quantitative difference likely arises from several factors. First, PSO represents a 6-month seasonal average of MDA8 O₃, while ODVs target specific high-percentile episodic conditions, and direct contributions to seasonal averages can be expected to differ from enhancements during specific episodes (although episodic contributions could be expected to have a higher share of local photochemistry than seasonal contributions). Second, and perhaps more fundamentally, inferential methods based on subtracting an estimated 'baseline' from total observed ozone may systematically underestimate the full impact of local anthropogenic emissions. Such approaches often define the baseline based on remote sites or specific statistical filtering, which may not fully account for the ozone produced from local emissions that is then regionally dispersed (as we also see indications of anthropogenic NO_x and BVOC interactions in the tagged output) or the non-linear chemical feedbacks that occur when local emissions are present. In contrast, our emissions tagging technique directly attributes ozone formation to its original precursor sources as they undergo transport and chemical transformation within the model's complete and consistent chemical framework. This provides a mechanistic quantification of source contributions to the specific PSO metric under baseline conditions. While inferential methods provide valuable observational constraints, our tagging approach offers a complementary, process-explicit view of how different source categories contribute to the ozone burden, particularly illuminating the partitioning between local, regional, and intercontinental sources in the complex, evolving atmospheric environment”

Multiple other modeling studies have reported contributions to ozone that differ quantitatively from the present results. For example, Mathur et al. (2022b) find that “stratospheric O₃ (ranging between 6 and 20 ppb) constitutes 29%–78% of the estimated Spring-time background O₃ across the continental United States” while the present paper quantifies significantly smaller impacts: “the stratosphere contributes up to 6–8 ppb in the Southwestern US” (the US region of maximum stratospheric influence) and 4–7 ppb in the Northeastern US. Comparisons and contrasts of quantitative estimates from multiple studies are required in this paper.

We have now included comparisons with reported contributions from published literature.

L979-986: “The springtime (March-May) ozone has seen increases in both foreign NO_x contributions (13.16 ppb to 14.81 ppb) as well as stratospheric contributions (12.02 ppb to 12.55 ppb; see Table 3 for a comparison across regions). Springtime mean stratospheric contribution is 12.55

ppb in the recent period (even higher in SW US at 14.25 ppb; Figure S7; Table 3). Previous studies have reported modelled stratospheric contributions in North America during observationally-identified episodes with higher values (e.g., 20–40 ppb; Lin et al., 2012) as well as seasonal mean contributions (6–18 ppb; Mathur et al. 2022b). Our seasonal mean values are lower likely because we do not sample the model output extensively from the mountainous region of western US, where stratospheric contributions are highest, due to lack of TOAR observations in those regions.”

A clear and concise summary of what new understanding of the atmosphere has emerged from this study is lacking. The final section of the paper discusses Conclusions, Limitations and Future Outlook; it lists many findings, but it is not clear to me either what is new in this analysis, or which of the findings required the dual tagging system to uncover. That material should be revised to clearly and specifically answer several questions: What new knowledge of atmospheric chemistry emerged from this work? The paper does utilize a relatively novel tagging approach; can the authors provide the reader with a concise summary of when or for what issues the joint NO_x and VOC source tagging is required? Or can they at least clearly summarize what additional information was provided by that technique in this study? (After all, the technique does greatly complicate the analysis, and in the end the added benefits are not clear to me.)

We have now significantly updated the Conclusions section to emphasize the new knowledge emerging from this study (L1385–1426) and the role of the TOAST 1.0 dual-tagging system in producing such knowledge (L1431–1445).

Major issues:

The format of Figures 5 and 9 should be improved to better illustrate the authors' discussion. Most of the source contributions are so small that their magnitudes and variation are difficult to discern in the present format. Improvements should include a) using a more nearly square format to more clearly show any systematic changes over the two decades, and b) perhaps using a log-scale for the ordinate to more clearly illustrate the magnitude of all source contributions. The nearly square format would be more easily obtained by a) moving the region labels into blank spaces within the graphs and b) by not repeating the years on the abscissa of each graph. (Similar comments apply to Figures 7 and 10.) The log-scale would also be more appropriate for showing the long-term ozone changes due to changes in anthropogenic emissions, since those emissions are expected to decrease in an approximately logarithmic fashion (i.e., linear on the log scale). Figures 6, 8, 11 and 12 are more readable, but could be improved by changing the ordinate scale to 0 to 30 in the 2nd and 3rd column graphs; this would cut off the top of the anthropogenic NO_x contribution for the NE US, but it would be useful to duplicate this contribution in all of the 1st column graphs.

We have made several changes to the PSO time series figures for both NAM and EUR (now figures 6 and 12) to address the concerns of the reviewer. First, the NO_x and VOC contributions are plotted on separate panels from the total PSO. So, there are 3 columns in the new panelplot. The height:width ratio of the panels has been increased and the maximum ordinate on the y-axis

has been cut off at 40 ppb for NAM and 30 ppb for EUR for the PSO contributions panels to ensure better discernibility of the trends for individual contributions. We have retained the linear-scale instead of the log-scale in line with our decision to fit linear trends to these contributions.

We have also modified the old seasonal cycle figures for NAM and EUR (figures 6 and 10). The height:width aspect ratio has been increased in the new figures (Figures 8 and 14) for improved visibility of the smaller contributions.

For the figures showing seasonal cycle change in NAM and EUR (now fig 9, 10, 15, 16, S6-S10), we have cut off the maximum ordinate on the y-axis for the middle and right panels (those showing source contributions) to 40 ppb to ensure better visibility. For NE US, as recommended by the reviewer, we have duplicated the local anthropogenic NO_x contribution to the average seasonal cycle over 2000-2004 on the main panel with total MDA8 O₃ seasonal cycle for that region (Figure 10).

In Section 3 the authors discuss the long-term changes in ozone and its components in terms of linear trends derived from a Theil-Sen approach; these trends are collected in Table S1. Several issues should be discussed in this regard. First, the emissions illustrated in Figure 2 appear to be non-linear; thus, at least some of the long-term changes can be expected to be non-linear. The authors should discuss why they employ a technique that can only quantify the linear aspects of the long-term changes. Logan et al. (2012) quantify changes in linear slopes (i.e., trends) over a 3 decade long data record; perhaps such an analysis should be employed in the present discussion?

Logan et al. (2012) and others (e.g., Parrish et al., 2025; Parrish et al. 2020) indeed chose to fit a quadratic function to ozone trends in North America and Europe. We believe that this choice was appropriate given the longer time-series considered in these studies which included both an initial increase in observed ozone in the 1980s and 1990s followed by a decrease in the 2000s. However, for our study, we focus on the 2000-2018 period when the ozone trends in these regions have consistently declined and, based on a visual inspection, concluded that linear trends (along with p-values and 95% CI) as recommended in TOAR statistical guidelines (https://igacproject.org/sites/default/files/2023-04/STAT_recommendations_TOAR_analyses_0.pdf) are appropriate. We have now mentioned this point in the manuscript:

L538-545: “Figure 6 presents the time series of observed and model-simulated total PSO (panels a, d, g, j, m), alongside the attributed contributions from NO_x sources (panels b, e, h, k, n) and VOC sources (panels c, f, i, l, o). On a visual inspection of observed and modelled PSO trends (left column panels) we decided to fit Generalized Least Squares (GLS) linear trends to these data points. We note that some previous studies have fitted higher order functions to ozone data over North America as necessitated by their longer period of analysis where ozone concentrations increased, stagnated, and then decreased (Logan et al., 2012; Parrish et al., 2025; Parrish et al. 2020). However, a linear fit is appropriate for the period considered in our study when local emissions have only declined (Figure 2). Quantitative details of the trends and their significance for all contributions are provided in Table 2.”

Second, the authors report values they derive for the significance of their derived trends; this significance only informs us regarding whether the trends are significantly different from zero (it

is not clear to me what is implied by a trend of zero with a significance of 1 – that seems nonsensical). The authors should report 95% confidence limits for their derived trends if they indeed judge a linear analysis is adequate to quantify statistically significant long-term changes; these confidence limits are of much greater interest than the significance statistic, as they provide a basis for judging quantitative comparisons such as the authors give on lines 451-453; as presently written, it is not clear that the -0.24 ppb/yr (1.0) trend derived from the observations differs significantly from the -0.35 ppb/yr (0.99) trend derived from the model results.

We have now fitted linear trends based on Generalized Least Squares (GLS) method instead of the Theil-Sen method, as recommended by the TOAR statistical guidelines, and have also included p-values along with 95% confidence intervals. These are included in Table 2.

Finally, and most importantly, the significance values (and potentially any calculated confidence limits) are apparently greatly over-optimistic. If I understand correctly, each region has only a single PSO value each year. Given the limited (i.e., 19) number of PSO values combined with the relatively large interannual variability and autocorrelation that characterize observed time series of ozone concentrations, only modest significance values (and relatively wide calculated confidence limits) are expected. Importantly, for time series of annual PSO values, autocorrelation over multiple year must be considered in deriving reliable confidence limits. Fiore et al. (2022) discuss this issue more fully.

The new GLS linear trend analysis on the new data (with updated TOAR-II data and model co-sampling) yields lower significance values (i.e., higher p-values) and broader 95% confidence intervals. Autocorrelation was considered while deriving these trends and confidence limits.

More generally, the authors provide few confidence limits for the quantitative numbers given. It is generally acknowledged that any scientific paper presenting results of quantitative analysis must include confidence limits for the quantitative findings. This last comment applies to all of the quantitative results presented in the paper. I realize that it is difficult to quantify confidence limits for model results; nevertheless such quantification is essential. Developing such confidence limits could come both from the additional statistical analysis indicated as needed in the first Critical Issue discussed above, better quantitative treatment of the seasonal cycle and its shifts as suggested in the second Critical Issue, also discussed above, and further bolstered from quantitative comparisons of the results from the modeling study presented in this paper with other observational and modeling results, again indicated as needed in the second Critical Issue.

We have now provided 95% confidence limits for both observed and modelled PSO as well as modelled contributions to PSO (see Table 2). We have also taken a much more quantitative approach in model evaluation as well as analysis of seasonal cycles (sections 3.2.2, 3.2.3, 3.3.2, 3.3.3) by breaking down both observed and modelled seasonal cycles into two harmonics and comparing their various parameters, as discussed in previous responses.

The sentences on lines 496-499 are contradictory and obscure an important point: “The model reproduces the 19-year average seasonal cycle over different parts of North America *very well*. For western regions, we see a consistent systematic positive bias of 2-4 ppb. For eastern regions

we see a *very good reproduction* of the seasonal cycle during winter and spring but a *notable overestimation during summertime* (italics added).” The italicized words are where the contradiction arises. The authors further discuss the summertime overestimate, in multiple places. I suggest a single, consistent discussion of this feature over all of the North American regions.

We have now rewritten this section with a quantitative description and the supposedly contradictory sounding sentences have been removed.

Lines 532-533 state that “Figures 6e and f show the average seasonal cycle of MDA8 O₃ in Southwestern US which is similar to that for the Northwestern US” I agree that the shapes are somewhat similar between the two regions, but the minimum-maximum difference is significantly smaller in the Southwestern US (~12 ppb) than in the Northwestern US (~22 ppb). This mis-judgement, evidently based on a qualitative assessment of Figure 6, emphasizes the need for the utilization of a quantitative analysis of the ozone seasonal cycles, as detailed above in the discussion of second Critical Issue. There is a similar but smaller mis-judgement in the following comparison of the Southeastern US and Northeastern US seasonal cycles.

These seasonal cycles are now discussed in terms of their Fourier parameters which avoids such visual mis-judgement.

Minor issues:

Line 31: The phrase “... especially towards the end of the 20th century” would be more accurate if changed to “... especially during the last half of the 20th century”. Substantial ozone increases in the troposphere have been documented over that entire period.

We have now modified this sentence accordingly.

Line 56: The authors state that “This is due to the long-enough atmospheric lifetime of ozone (about 3-4 weeks) which allows it to traverse intercontinental distances and affect the air quality of regions far from the location of its chemical production or the location of the emission of its precursors.” This statement is true as written, but should be discussed a bit further. The loss processes leading to that lifetime are dominated by loss in warmer, more humid tropical regions. Further, this lifetime refers to the total photochemical loss processes integrated over the entire globe; considering net ozone tendency, the effective lifetime of ozone in an air parcel transported in the free troposphere at northern midlatitudes (the zone of focus of this paper) is on the order of several months. This is long enough that the free troposphere can be considered a reasonably well-mixed reservoir, further emphasizing the importance of transport over intercontinental distances within this latitude zone.

L62-70: We have now modified the text to the following: “This is due to the long-enough atmospheric lifetime of ozone which allows it to traverse intercontinental distances and affect the

air quality of regions far from the location of its chemical production or the location of the emission of its precursors. While the global average tropospheric lifetime of ozone is often cited as approximately 3–4 weeks, a figure largely influenced by more rapid photochemical loss in warmer, humid tropical regions (e.g., Stevenson et al., 2006; Young et al., 2013), the effective lifetime of ozone in air parcels transported within the cooler, drier free troposphere at northern midlatitudes is considerably longer, on the order of several months (e.g., Jacob, 1999; Wang and Jacob, 1998; Fiore et al., 2009). This extended lifetime in the primary transport pathway for intercontinental pollution allows ozone to traverse vast distances and enables the northern mid-latitude free troposphere to act as a relatively well-mixed reservoir (Parrish et al., 2020)."

Lines 63–64: The authors correctly note that studies have identified “increasing trends in wintertime and background ozone concentrations at many sites in North America, particularly at the US west coast”. For completeness, it would be useful to further point out that such increases have also been identified throughout the background troposphere at northern midlatitudes including in the free troposphere, but that ozone in this latitude zone reached a maximum in the first decade of the 2000s (e.g., Parrish et al., 2020; Derwent et al., 2024).

We have now added the following sentence in L79–81: “Such increases in ozone have also been identified throughout the background troposphere at northern midlatitudes including in the free troposphere, with a peak attained in the first decade of the 2000s (e.g., Parrish et al., 2020; Derwent et al., 2024)”

Lines 79–81: The authors could expand this sentence for completeness. As Derwent et al. (2024) discuss, a hierarchy of models is required to fully understand tropospheric ozone. We require not only statistical interpretations of observational data and well-evaluated atmospheric chemical transport models, but also conceptual models that simplify and capture the essence of the most salient physical and chemical processes that control observed ozone abundances.

We have now added the following text in L101–103: “Together, observational analyses and model-generated results can aid the theoretical development and improvement of simpler conceptual models that capture the essence of the most salient physical and chemical processes that control observed ozone abundances (Derwent et al., 2024).”

Lines 88–102: It seems to me that this paragraph overemphasizes the shortcomings of the *perturbation*. Is it not possible to simply limit the approach to such small perturbations that the atmosphere processes are not significantly changed, and the results approach perfect accuracy?

You raise a pertinent point regarding the theoretical limit of small perturbations. Indeed, for infinitesimally small changes in emissions where the atmospheric chemical regime remains essentially linear and unperturbed, the calculated sensitivity might more closely approximate a direct fractional contribution. However, practical source apportionment using the perturbation method typically involves substantial reductions (e.g., 20–100%) of an emission source or sector to obtain a clear and robust signal above model noise and inherent atmospheric variability. Such large perturbations inevitably alter the chemical regime (e.g., by changing NO_x/VOC ratios or HO_x cycling), meaning the difference between the perturbed run and the baseline reflects the impact of that source's removal rather than its contribution within the original atmosphere.

To make this clearer, we added an additional line in the text:

L124-126: *“On the other hand, tagging techniques, which track the fate of emissions from designated sources as they undergo transport and chemical transformation within the unperturbed baseline atmosphere, allow us to assess the contribution of various sources under a baseline scenario when no policy intervention has been made.”*

Table 1 lists 9 regional oceanic tagged regions. However, these regions are not mapped in either the paper or the Supplement. Please include such a map; for example, it is of interest to understand how the North Atlantic Ocean is divided into 3 separately tagged regions.

Thanks for pointing this out. We have now added a new global map of the tagged oceanic regions. See figure S12.

Lines 264-274 describe the derivation of MDA8 values from the model output, and lines 287-292 describe the derivation of MDA8 values from the TOAR rural observations. Please discuss if these methods are completely compatible, or if differences in the procedures may possibly be important.

The procedure for calculating MDA8 O3 is essentially the same for both observations and model output. The fact that MDA8 O3 is only calculated for days where at least 18 of the 24 hourly values are available in the observations allows us to minimize any discrepancies between the observed and model-derived MDA8 O3. In other words, the model-derived MDA8 O3 is only sampled into further analyses (i.e. spatial averaging over receptor regions) if the corresponding TOAR-MDA8 O3 has >18 hourly data points. We have emphasized this in the text by adding a line: *“This allows us to minimize any discrepancies between the observed and model-derived MDA8 O3 values”*. (L342-343)

Lines 278-279 state: “We use these receptor regions to perform area-weighted spatial averaging of MDA8 O3 values before analysing the trends and contributions. Please explain the process of “area-weighted spatial averaging”, and why it is used rather than simple averaging.

We have now added the following text in L323-330: *“Area-weighted spatial averaging is needed because different model grid cells cover different areas on the ground based on the rectangular lat-long coordinate system, with high-latitude grid cells covering smaller areas and low-latitude and equatorial grid cells covering larger areas. So, a simple spatial averaging will overrepresent the concentrations of high-latitude gridcells and underrepresent lower-latitude gridcell concentrations in the receptor region average. So, we derive dimensionless coefficients ranging for all grid cells within each receptor region based on their relative size to the average grid cell area in that region. We scale the gridded MDA8 O3 with these area-coefficients before averaging, ensuring a proportionate representation of the MDA8 O3 value over the entire receptor region.”*

Lines 405-406 note “that for all regions in North America, the observed PSO exceeds the WHO guidelines throughout the 2000 -2018 period.” It should be emphasized that the WHO guideline is based on the highest tail of the MDA8 distribution, while mean values are discussed in this

paper. Thus, the exceedance of the WHO guideline over North America is indeed profound, as reflected in the difference in the WHO (~50 ppb) and US EPA (70 ppb) guidelines.

We do not analyze daily exceedances in our study but focus on Peak Season Ozone (PSO) and the WHO guideline for PSO (~31 ppb) is based on a maximum 6-month *average* MDA8 O₃ which is exactly what we use in this study.

Line 450 contains a typo – the WHO guideline is 51 ppb.

By the WHO guideline, here, we are referring to the PSO guideline by the WHO, which is indeed 60 ug/m³ or ~31 ppb. More details are available on page 102 of this report:

<https://iris.who.int/bitstream/handle/10665/345329/9789240034228-eng.pdf>

Each panel of Figures 5 and 9 illustrates a time series of a “Residual O₃” contribution. I have not found that quantity defined or discussed in the manuscript. Can this be eliminated? If not please define and discuss.

Thanks for pointing this out. This “residual ozone” is ozone which can not be clearly attributed to either a NO_x or a VOC precursor, and which is not associated with the photolysis of molecular oxygen in the stratosphere,, for example the ozone formed from O atoms resulting from the self-reaction of hydroxyl radicals. We have added a sentence to the discussion of the tagging system in the methodology section (lines 239-243) describing this. This residual ozone can not be eliminated, but typically only makes a very small contribution to ozone mixing ratio at the surface (about 1 ppb in this study).

Lines 727-728 state that “the increasing contribution of natural NO_x emissions we find in our study, especially during the summertime, is most likely due to the increasing ozone productivity of these emissions.” This appears to be speculative; if this statement is to be included in the Conclusion section, it should be shown to be true through quantitative analysis.

L1394-1397: We have added further reasoning to the text to support the statement: “*the increasing trend in modelled contribution of natural NO_x emissions, especially during the summertime, suggests increasing ozone productivity of these emissions since there is no noticeable increasing trend in natural NO_x emissions and a slight decreasing trend in Lightning NO_x emissions (Figure 3 a, c, e, g).*”

Lines 729-730 state that there is “... a smaller effect in the springtime, when long-range transport of ozone produced from foreign anthropogenic NO_x emissions is more important.” I believe that in the literature there is ongoing discussion regarding whether ozone produced from foreign anthropogenic NO_x emissions or ozone of stratospheric origin is of most importance for springtime surface ozone over North America (if not also Europe); I suggest that the authors mention both of these sources in this context.

Thank you for this important point. We have now modified the text as “*when long-range transport of ozone produced from foreign anthropogenic NO_x emissions and stratosphere is more important*”. We have also now quantified the springtime foreign NO_x and stratospheric contributions to all regions in Table 3.

Community Comment (by Owen Cooper)

General comments:

There are some data issues that need to be addressed:

1) Figure S8 (top panel). Please check the North American data availability for 2012. This figure indicates that very few sites are available, but I went to the TOAR-II database and using the plotting tool I found plenty of data for that year. Please see the example figure pasted below.

We have updated the TOAR data (please also see the response to reviewer 2 in this regard) and now there is no significant drop in the no. of North American stations in 2012. All model data used in subsequent analyses is also updated based on new co-sampling corresponding to the updated TOAR-II data.

2) Line 370. Here you mention that you compared the model to observations in regions with limited data. But are there enough data available for a meaningful comparison? For example, the TOAR map tool doesn't show any data in Belarus or Ukraine, and only urban data in Russia. The same applies for North Africa. Please indicate the number of rural stations with ozone data in all of the regions analyzed in this study, and explain how representative you believe the data to be.

We have now updated the text with the number of TOAR stations as available in the TOAR-II gridded dataset. The model-observation representativeness is maintained by co-sampling the model output from only those gridcells where TOAR stations are present. We have highlighted this in the updated text: *"We have also included the Belarus & Ukraine region (Figure 4j; with 1-2 valid stations) in our evaluation and here too we see a good simulation of MDA8 O₃ for the entire period (with a small mean bias of 0.56 ppb and r value of 0.83), barring a couple of years (2014 and 2017) when the model overestimates the values. We have also evaluated the model for MDA8 O₃ against rural observations from the TOAR-II database in other regions including Mexico (11-14 stations), North Africa (1-3 stations), Southern Africa (1 station), Southern Latin America (1-2 stations), and European Russia (2 stations; see Figure 3 for region definitions), where the model has also captured the trends well, however, since we do not discuss these regions in further analyses, they are presented in the supplement (see, Figure S1). Here too, the model output is extracted only from those grid cells where at least one TOAR station exists, ensuring representative co-sampling."*

In the introduction it would be helpful to cite some of the key papers from the first phase of TOAR as they are highly relevant to the background information on the importance of ozone for health, vegetation and climate: Fleming and Doherty et al. (2018), Mills et al. (2018), Gaudel et al. (2018). Another key reference for ozone's impact on climate is IPCC AR6 (Szopa et al., 2021).

Fleming and Doherty et al. (2018), Mills et al. (2018), and (Szopa et al., 2021) have been included in the text with appropriate context. See lines 47-52.

Lines 63-67

Regarding the observed increase of background ozone at northern mid-latitudes, some of the cited references are out of date and other recent studies have not been cited. IPCC AR6 (Gulev et al., 2021) assessed an increase of free-tropospheric ozone at northern mid-latitudes through the year 2016. Follow-up studies by Wang et al. (2022) and Christiansen et al. (2022) used the updated IAGOS and ozonesonde records, and showed similar results to IPCC AR6. New papers that have emerged from the TOAR-II effort show ozone increases through 2019, especially over East Asia (Eshorbany et al., 2024; Lu et al., 2024). It is also worth mentioning several recent studies that show ozone increased through at least 2019, but since 2020 ozone has decreased slightly, or levelled off, in association with the COVID-19 economic downturn (Miyazaki et al., 2021; Chang et al., 2022; 2023; Ziemke et al., 2022). A good review of U.S. background ozone is provided by Jaffe et al., 2018.

We thank the reviewer for providing this comprehensive list of recent papers with related research. On a careful reading of the above papers, we found the findings of Jaffe et al., 2018; Christiansen et al., 2022; Elshorbany et al. 2024; and Lu et al., 2024 to be directly relevant to the discussion in our study and have included them in the text now. Christiansen et al., 2022 has been added to the list of citations in L79, and the following line is added mentioning Jaffe et al., 2018: *“Jaffe et al., (2018) performed a comprehensive knowledge assessment of background ozone in the US and emphasized its growing relative importance and advocated for, among other things, a more strategic observational network and new process-based modelling studies to better quantify background ozone in the US to support informed clean air policies.”*

Elshorbany et al., 2024 and Lu et al., 2024 are included later in the results section (section 3.4) *“These results are consistent with findings of Elshorbany et al., (2024) and Lu et al., (2024) who report increasing ozone trends in Asia both in the troposphere and at the surface which stabilize around 2013”.*

Table S1, and wherever trends are reported in the paper

As described in the TOAR-II Recommendations for Statistical Analyses, all trends need to be reported with 95% confidence intervals and p-values. It's not clear what is meant by “and their significance (shown in brackets).”

All trends are now reported along with p-values in parentheses and 95% confidence intervals in square brackets at all instances in the text as well as in Table S1. The trends were also revised based on Generalized Least Squares (GLS) fit in line with the TOAR statistical guidelines.

Line 349

For context, a very good review of ozone across the southwestern USA is provided by Sorooshian et al., 2024. Also, Simon et al., 2024 provide an update on current impacts of emissions on U.S. surface ozone.

Since we have substantially rewritten this section, this citation didn't fit well at this place into the new text, however, we have included many references from within Sorooshian et al., 2024 (which is indeed a rich review) in the results section (section 3.2.3; L975-977).

We have also cited Simon et al. (2024) in section 3.2.1 L563-566: “These results are consistent with findings from Simon et al. (2024) who analysed observational trends over 51 sites in the US over roughly the same period (2002-2019) and found the marked impact of clean air policies across the US such that the difference between the weekend (lower NO_x) and weekday (higher NO_x) MDA8 O₃ has diminished and become negative in recent years reflecting a transition from NO_x-saturated to NO_x-limited ozone formation regime.”

and in L575-577: “This lack of correlation between local NO_x emissions and observed MDA8 O₃ has been reported by Simon et al. (2023) for rural California even at a higher temporal frequency through disappearing day-of-week activity patterns indicating an increasing role of transported ozone in this region.”

Line 393

It would be helpful to briefly mention how your PSO values compare to previous studies, such as Becker et al., 2023.

We have now included a reference to Becker et al. (2023) when discussing spatial PSO results for NAM and EUR (these are new maps in the revised manuscript).

L531: “The spatial features of PSO for both years are very similar to bias-corrected maps of PSO for 2000 and 2017 presented in Becker et al. (2023).”

And L1113-1116

“However, this decline is not obviously reflected in the spatial patterns of total PSO (Figure 11b, e), which generally decreased in the southern regions but not in northern regions, especially over areas with the largest emission cuts, as also seen in bias-corrected PSO maps by Becker et al. (2023)”

Line 594

It's worth pointing out that the change in ozone seasonal cycle is expected to continue to shift with future emissions changes, as discussed by Clifton et al. (2014).

We have included this reference and included an extra line “This transition in the ozone seasonal cycle in the NE US, towards a springtime maximum, is expected to continue with future emissions changes, as discussed by Clifton et al. (2014)” in section 3.2.3 at L1032.

The submitted paper provides few references regarding the important role that methane plays in ozone production. A recent assessment of the impact of methane on tropospheric ozone is provided by the UN Climate and Clean Air Coalition:

United Nations Environment Programme and Climate and Clean Air Coalition (2021). Global Methane Assessment: Benefits and Costs of Mitigating Methane Emissions. Nairobi: United Nations Environment Programme. ISBN: 978-92-807-3854-4. Job No: DTI/2352/PA
https://www.ccacoalition.org/sites/default/files/resources//2021_GlobalMethane_Assessment_full_0.pdf

We have now included this reference in the introduction at L42.