We would like to thank the reviewers for their comments. Below are responses to the individual comments. Responses are in blue, and revised text in the manuscript is in red.

Reviewer 1:

Background O_3 constitutes a significant portion of total surface O_3 and the contribution becomes higher when O_3 levels decline. This study used a measurement-model data fusion approach to assess CTM biases in USB O_3 and attribute these biases to different sources. Two sets of CMAQ simulations, PA and EQUATES, are conducted to analyze the contributions of different sources to USB O_3. While the study is of scientific significance in assessing model biases in background O_3 estimation, two major concerns need to be addressed by the authors.

(1) The two sets of simulations are for different years, the model configuration and inputs are different, different versions of emissions inventory are adopted, which will introduce substantial uncertainties. The PA simulations cover the year of 2016; while the EQUATES simulations span between 2002-2019. Additional simulations using the EQUATES modeling framework were conducted for 2016–2017 to estimate USB O_3 and USA O_3 using the zero-out method. CMAQ v5.2.1 was used for the PA simulations while CMAQ v5.3.2 was used for the EQUATES simulations. It seems this study is not outlined within a comprehensive framework but combine different modelling work to do the current study. The differences of biases caused by different model configurations and model setup between the two sets of scenarios need to be fully discussed. (2) While biases in the study are fully discussed, the reasons behind these biases remain ambiguous. Factors such as uncertainties in emissions inventory, meteorology simulations, and chemical mechanisms could contribute to biases. Providing insights into the main drivers of biases and offering suggestions for modelers to mitigate these biases would enhance the value of the study for readers seeking to improve background O_3 modeling accuracy.

(1) It is true that the Policy Assessment (PA) and EQUATES simulations were not conducted specifically for this study. However, an opportunity to use these datasets for the analysis presented here was available. Some major differences between the two sets of simulations that are expected to affect O_3 are given in the final paragraph of the original introduction. One is the addition of halogen chemistry which was added to CMAQ between the model version used for the PA simulations and the model version used for the EQUATES simulations. The second is a different US anthropogenic emissions inventory. Further details on model configuration of both sets of simulations are given in Tables S4-S5. See the response to comment below (2. Methods) for additional details that have been added to describe the model configuration.

(2) Agreed that it would be great if we were able to identify the specific reasons for biases here, but we are not able to assess that with confidence in the current study. This work is intended to be largely descriptive rather than proscriptive. The following additional discussion has been added to Section 4 to clarify and expand on this:

“While details on the spatial and temporal characteristics of biases in different O_3 components are provided here, the correlational bias attribution method employed here does not necessarily identify the specific factors that drive the biases. These results provide estimates of potential biases in USB and USA O_3 that can inform more targeted future work examining the individual sources in greater detail.”

Specific comments:

1. **Abstract**. The current form of the abstract is notably objective, it lacks quantitative results detailing the biases and their spatial-temporal characteristics. Additionally, what can we do to reduce these biases? The abstract could benefit from discussing potential strategies to mitigate these biases.

We have added the following quantitative results to the abstract:

"Summer average US anthropogenic O_3 in the eastern US was estimated to be biased high by 2, 7, and 11 ppb (11%, 32%, and 49%) for one set of simulations at 12, 36, and 108 km resolutions and 1 and 6 ppb (10% and 37%) for another set of simulations at 12 and 108 km resolutions."

and

"Despite this, results indicate a negative bias in modeled estimates of the impact of stratospheric O_3 at the surface, with a western US spring average bias of -3.5 ppb (-25%) estimated based on a stratospheric O_3 tracer."

See item (2) of the previous comment for a response on the discussion of potential strategies to mitigate the biases.

2. **Methods**. The overall model configuration needs to be briefly outlined in the main text, such as modelling domain, WRF configuration, CMAQ gas-phase mechanism, IC/BC, vertical layers, etc.

Many of these details are available from Tables S4 and S5. Information on the vertical layer structure and more details on the modeling domains have now been added to Table S5. We have also added additional details given below in Section 2.1 of the manuscript. (Note this response is identical to the response to reviewer #2 comment marked "Table 1")

Vertical structure:

"Both the PA and EQUATES simulations use a 44-layer vertical structure for hemispheric scale applications (at 108 km resolution) and a 35-layer vertical structure for continental (i.e., 36 km and 12 km resolution) applications with a vertical extent from the surface to 50 hPa and a surface layer height of approximately 20 m for both the hemispheric and continental configurations (see Mathur et al. (2017) for more details on these vertical layer structures)."

Model configuration (chemical mechanism and meteorological model version):

"Besides the addition of halogen chemistry, there are other differences in the chemical mechanisms used for each set of simulations. The mechanisms used for the hemispheric simulations were cb6r3_ae6_aq for the PA simulations and cb6r3m_ae7_kmtbr for the EQUATES simulations. The part of the mechanism name labeled cb6r3m indicates additional chemistry relevant in marine environments (the halogen chemistry described above); ae6 and ae7 indicate the version number for chemistry relevant to aerosols;

aq and kmtbr indicate different treatments of cloud chemistry. The chemical mechanisms used for continental-scale PA and EQUATES simulations (cb6r3_ae6nvPOA_aq and cb6r3_ae7_aq) also differ in their representation of organic aerosols (Murphy et al., 2017; Pye et al., 2019; Qin et al., 2021; Appel et al., 2021) which could affect O_3 concentrations. Different versions of WRF (v3.8 for PA simulations and v4.1.1 for EQUATES simulations) employed may also contribute to differences in O_3."

Emissions and stratospheric O_3:

"Emission inputs also differ between the PA and EQUATES simulations. Different US anthropogenic emission inventories were used for the simulations. The PA simulations used an early version (sometimes called the "alpha" version) of a 2016 emissions modeling platform developed by the National Emissions Inventory Collaborative (US EPA, 2019a). The EQUATES simulations used an inventory that was developed as part of the broader EQUATES framework to model a long timeseries using consistent methods for emissions estimates (Foley et al., 2023). For emissions in Canada and Mexico, both sets of simulations use emission inventories developed by the respective national governments, though the EQUATES simulations use more recent inventories (as described by Foley et al. (2020)) than the PA simulations (as described by US EPA (2019a)). Both the PA and EQUATES simulations use the Tsinghua University inventory of emissions in China (Zhao et al., 2018). For other countries, both sets of simulations use the Hemispheric Transport of Air Pollution (HTAP) v2.2 inventory (Janssens-Maenhout et al., 2015) with scaling factors derived from the Community Emissions Data System (CEDS) (Hoesly et al., 2018) to account for yearly changes. Differences in the anthropogenic emissions used in the two model configurations are expected to contribute to differences in simulated O_3, most notably for the different US anthropogenic emissions since we focus here on O_3 in the US.

For hemispheric-scale simulations, biogenic VOC emissions are from the Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1) (Guenther et al., 2012). The PA simulations additionally replace MEGAN emissions with emissions from the Biogenic Emission Inventory System (BEIS) (Bash et al., 2016) over North America (USEPA, 2019b). The EQUATES MEGAN emissions are obtained from a compilation by Sindelarova et al. (2014). Soil NO_x emissions for the PA hemispheric simulations are also from MEGAN with replacement by BEIS soil NO_x over North America. Soil NOx emissions for the hemispheric EQUATES simulations are from a dataset by the Copernicus Atmosphere Monitoring Service (CAMS, 2018) based on methods by Yienger and Levy (1995). Lightning NO emissions for both the PA and EQUATES hemispheric simulations are from monthly climatology obtained from the Global Emissions Initiative (GEIA) and are based on Price et al. (1997). Lightning NO_x was not included in the PA continental-scale simulations, while lightning NO_x for the EQUATES continental-scale simulations is calculated using an inline module in CMAQ (Kang et al., 2019). For both PA and EQUATES, wildfire emissions outside of North America are based on the Fire Inventory from NCAR (FINN) v1.5 (Wiedinmyer et al., 2011) which provides day-specific fire emissions. Wildfires are vertically allocated with 25% of emissions distributed to the lowest two layers (~0-45 m), 35% distributed to layers 3-9 (~45-350 m), and the remaining 40% distributed to layers 10-19 (~350-2000 m) as described in the Technical Support Document for northern hemispheric emissions (US EPA, 2019b). Wildfire emissions within North America are based on the Hazard Mapping System (HMS) fire product which provides day-specific fire activity data. Emission processing for North American wildfires is further described in the Technical Support Document for North American emissions (US EPA, 2019a) (applicable to PA simulations) and Foley et al. (2023) (applicable to EQUATES simulations). Although the methods are similar, North American wildfire emissions may differ between PA and EQUATES based on the specific fire activity data that was used in each case. Fire plume injection height for North American fires is determined by an inline plume rise algorithm in CMAQ based on fire heat content (see e.g., Wilkins et al. (2022) for more details on fire plume injection height in CMAQ). Stratospheric O_3 in both the PA and EQUATES simulations is from the PV parameterization by Xing et al. (2016) (described in more detail above) in the hemispheric simulations. Stratospheric O_3 in the continental-scale simulations only comes from any stratospheric O_3 inherited from the lateral boundary conditions provided by the hemispheric simulations."

3. **Table 1**. The descriptions provided in Table 1 could benefit from clarification regarding emissions from other regions. For instance, in the case of "ZROW," where all international anthropogenic emissions are removed, it's unclear whether this includes emissions from the United States, Canada, and Mexico. To enhance clarity, it is recommended to use a clear table format that separates regions with emissions using symbols such as "√" to indicate the presence of emissions and "×" to denote the absence of emissions. This approach will facilitate a more straightforward understanding of the emission scenarios across different regions. Additionally, **Table 1 and Table S1 is exactly the same**.

We have updated the description in Table 1 of the ZROW simulation to clarify:

Original: "All international anthropogenic emissions are removed including prescribed fires where possible."

Revised: "All anthropogenic emissions outside the US are removed including prescribed fires where possible (ROW = rest of world)."

We have also slightly updated the description in Table 1 of the ZANTH simulation to clarify that this represents zeroing out anthropogenic emissions globally and have added references to the simulation names given in Table 1 in Section 2.1 to further connect the text of this section to Table 1. With these clarifications, the suggestion of a new table with check and x markings should not be necessary.

It is intentional that Table 1 and S1 are the same. The reason is explained in the caption to Table S1 which is repeated below (relevant sentence in bold here for emphasis):

"Table S1. Simulation names and descriptions for hemispheric-scale and regional-scale simulations. Table adapted from 2020 O_3 Policy Assessment Table 2-1 (USEPA, 2020). **Table S1 is reproduced from Table 1 in the main text to aid in interpreting Tables S2 and S3.**"

1. In the main text, Tables S4 and S5 (Line 95) comes before Table S3 (Line 101), this is strange.

We have slightly rearranged the text to make the first mentions of Tables S3, S4, and S5 appear in order.

2. Line 115, "STRAT" should be spelled out at the first time it appears.

We have made a small change to the relevant sentence to define the meaning of STRAT.

Original: "We refer to the PV tracer concentrations as STRAT O_3 since it relates to the stratospheric influence, but it only partly replicates the impact of stratospheric O3 since it does not undergo chemical losses."

Revised: "We refer to the PV tracer concentrations as STRAT (short for stratospheric) O_3 since it relates to the stratospheric influence, but it only partly replicates the impact of stratospheric O3 since it does not undergo chemical losses."

3. **CTM results**: can you explain further why the 12km simulations have the best performance? Additionally, what about the model performance for NO2 simulations?

The original statement in the manuscript that the 12 km simulations have the best performance as indicated by the normalized mean bias is incorrect, as the 108 km EQUATES simulations have normalized mean bias (NMB) closer to zero compared to the 12 km EQUATES simulations. This sentence has therefore been removed, and we now simply present the performance for each simulation as summarized by the NMB. The role of different O_3 components in these biases is further discussed in Sections 3.3 and 3.4.

While the performance for NO_x is of course relevant for O_3, a full performance evaluation for NO_2 or NO_x is considered beyond the scope of the current paper.

4. Table 2 and 3, what's the unit of the data in this table?

The following has been added to the captions of Tables 2 and 3:

"Numbers in the table are in units of ppb."

5. **section 3.3.** The authors extensively compared the differences in deviations between different scenarios. Are these differences in deviations related to the zero-out method neglecting nonlinear ozone production?

It is not clear that the nonlinear ozone production which is not accounted for in the zero-out method would result in differences in the inferred biases between the PA and EQUATES modeling setups. Both setups use the zero-out method to estimate US background and US anthropogenic contributions. We have added some additional details in Section 2.1 to describe how the zero-out method neglects non-linear interactions between different emissions sources and why it is preferred to other techniques here:

"The zero-out method is the most common approach for simulating USB O_3, though other approaches such as sensitivity simulations and source tagging techniques have also been previously employed (Jaffe et al., 2018). The zero-out method neglects non-linear interactions between sources which can affect the simulated source contribution (Wu et al., 2009; Dolwick et al., 2015). However, the zero-out method is consistent with the definition of USB O_3 as the level of O_3 in the absence of US anthropogenic emissions, while sensitivity or tagging techniques would instead provide an estimate of source contributions to total simulated O_3 (including O_3 from US anthropogenic sources)."

6. **section 3.4.** The authors extensively described the results of model bias. These results seem to be an extension of section 3.3, but what can these results further illustrate?

The following additional motivation for Section 3.4 has been added to the beginning of Section 3.4:

"The contributions and biases of different O_3 components have been presented so far as annual or seasonal averages (Figures 2-6 and 8) or as daily averages over US model grid cells (Figures 7 and 9). However, the relative contributions of O_3 components at different total O_3 concentrations is also of interest. For example, the relative contribution of USA and USB O_3 to total O_3 may be different on days with higher total O_3 vs. days with lower total O_3."

Additionally, this section focuses on the impacts at ozone monitoring sites, which are relevant for regulatory purposes. This is already stated in Section 3.4 so no additional text is needed to describe this.

Reviewer 2:

Major comments:

1. This study analyzed surface ozone from a suite of hemispheric and regional-scale CMAQ simulations for 2016 and 2017 and attempted to attribute the biases in model simulated total surface ozone to different components, including ozone produced from US anthropogenic emissions, natural sources, intercontinental transport, and stratospheric intrusions. Understanding US background ozone and its components is of broad interest because they are directly relevant to the setting and implementation of US ozone air quality standards. However, the manuscript needs to be substantially revised before it can be published. Description of the methodology used and discussions in many sections are incomplete. The authors should also discuss the model biases in the context of published literature. The referee's main concern is on the methodology used to attribute the model biases to different components. The description of the data fusion model in Section 2.3 is hard to understand. Is the data fusion model trained using one set of simulations and applied to another set of simulations for the bias attribution? How do you know the sources of biases in the two sets of simulations are the same? There are a couple of places where the authors refer to Skipper et al. (ES & T, 2021) for the method, but that study did not discuss the different USB components.

The description of the data fusion model approach in Section 2.3 has been reorganized and expanded as shown below to clarify:

Original: "Each O_3 component is multiplied by the alpha adjustment factor which varies as a function of space and time. The longitude and latitude terms are intended to capture the spatial variability of O_3 biases while the z term is intended to capture biases in O_3 related to elevation. The sinusoidal day of year terms are intended to capture the cyclical nature of O_3 production and to identify any seasonal dependence in O_3 biases."

Revised: "Each simulated O_3 component ($O3_i^{simulated}$) is multiplied by the alpha adjustment factor for that component (α_i), which varies as a function of space and time, to calculate an adjusted estimate of each O_3 component. The inferred model bias for a particular component is calculated as the difference between the original simulated O_3 and adjusted O_3 for that component. The individual adjusted O_3 components are summed to calculate the total adjusted O_3. The longitude and latitude terms of α_i are intended to capture the spatial variability of O_3 biases while the z term of α_i is intended to capture biases in O_3 related to elevation. The sinusoidal day of year terms of α_i are intended to capture the cyclical nature of O_3 production and to identify any seasonal dependence in O_3 biases."

Regarding the questions "Is the data fusion model trained using one set of simulations and applied to another set of simulations for the bias attribution? How do you know the sources of biases in the two sets of simulations are the same?":

A separate data fusion model is developed for each individual model configuration. Each model is only applied to the particular model configuration that it was trained on. We have updated part of Section 2.3 to clarify this:

Original: "A separate regression model is developed for each model resolution and USB O_3 component split. There are three model resolutions and three USB O_3 splits for the PA simulations, resulting in nine

PA models. There are two model resolutions for the EQUATES simulations. The 12 km EQUATES data has two USB O_3 splits while the 108 km EQUATES data has one USB O_3 split, resulting in three EQUATES models. For the PA models, only 2016 data is used since these simulations are for only that year. The models are trained on both 2016 and 2017 data for the EQUATES data."

Revised: "A separate regression model is developed for each separate model configuration (i.e., model resolution, PA or EQUATES simulation, and USB O_3 component split). There are three model resolutions and three USB O_3 splits for the PA simulations, resulting in nine PA models. There are two model resolutions for the EQUATES simulations. The 12 km EQUATES data has two USB O_3 splits while the 108 km EQUATES data has one USB O_3 split, resulting in three EQUATES models. For the PA models, only 2016 PA simulation data are used to train the models since these simulations are for only that year. For the EQUATES models, both 2016 and 2017 EQUATES simulation data are used to train the models."

The suggestion to add more discussion of published literature is addressed in subsequent responses.

2. The title of this paper is about the bias of US background ozone, but in the abstract and in the paper, there is substantial discussion on the biases of US anthropogenic O3 and the influence of model resolution. The authors stated "The estimated correction factors suggest a seasonally consistent positive bias in US anthropogenic O3 in the eastern US, with the bias becoming higher with coarser model resolution and with higher simulated total O3 though the bias does not increase much with higher observed O3." This statement seems to imply that coarser model resolution always produces higher US anthropogenic O3, which is not true. There is clearly a seasonal dependence. During winter when ozone production is in NOx-saturated regime, coarser model resolution leads to artificial dilution of NOx and thus higher O3 due to less NOx titration. During summer, however, when ozone production at most locations is in NOx-limited regime, coarser model resolution may lead to lower ozone concentrations produced from regional anthropogenic emissions. Increasing model resolution may lead to higher simulated US anthropogenic O3, leading to better agreement with observations, such as in the Central Valley of California. These seasonal characteristics of model resolution impacts on US anthropogenic ozone are clearly demonstrated in the published literature, including the recent studies of Schwantes et al. (2021) and Lin et a. (2024).

Schwantes, R. H., Lacey, F. G., Tilmes, S., Emmons, L. K., Lauritzen, P. H., Walters, S., et al. (2022). Evaluating the impact of chemical complexity and horizontal resolution on tropospheric ozone over the conterminous US with a global variable resolution chemistry model. *Journal of Advances in Modeling Earth Systems*, 14(6), e2021MS002889. https://doi.org/10.1029/2021MS002889

Lin, M., L. W. Horowitz, M. Zhao, L. Harris, P. Ginoux, J. P. Dunne, S. Malyshev, E. Shevliakova, H. Ahsan, S. Garner, F. Paulot, A. Pouyaei, S. J. Smith, Y. Xie, N. Zadeh, L. Zhou. *The GFDL Variable-Resolution Global Chemistry-Climate Model for Research at the Nexus of US Climate and Air Quality Extremes.* Journal of Advances in Modeling Earth Systems, in press, https://doi.org/10.1029/2023MS003984, 2024

The title has been updated to "Source specific bias correction of US background and anthropogenic ozone modeled in CMAQ" to reflect that there are findings relevant to both US background and anthropogenic O_3.

The finding of higher US anthropogenic O_3 biases at higher model resolutions applies to the eastern US, so the statement is not in conflict with the previous findings for O_3 simulated in the Central Valley of California. This statement is not meant to assert that coarser resolution leads to higher biases in all scenarios. It is only meant to apply to the specific findings here for the CMAQ simulations in the eastern US during warmer months. We have added some additional text to the abstract which clarifies the applicability of this finding:

"Summer average US anthropogenic O_3 in the eastern US was estimated to be biased high by 2, 7, and 11 ppb (11%, 32%, and 49%) for one set of simulations at 12, 36, and 108 km resolutions and 1 and 6 ppb (10% and 37%) for a second set of simulations at 12 and 108 km resolutions."

In Section 3.3, where similar results are discussed, we have added the following which includes discussion of findings in the recent articles mentioned by the reviewer:

"The inferred high biases in USA O_3 in the eastern US are primarily driven by biases in the summer and fall (Table S15, Figures S5-S7). Inferred eastern US USA O_3 biases average 2, 7, and 11 ppb in the summer and 3, 4, and 5 ppb in the fall for the 12, 36, and 108 km simulations. In the western US, where USA O_3 is mostly found to be biased low, coarser model resolution results in the summer average bias changing from slightly negative in the 12 km simulations (-0.5 ppb) to slightly positive in the 36 and 108 km simulations (+0.7 ppb and +1.0 ppb).

In contrast to our results showing an increase in O_3 with coarser resolution, Schwantes et al. (2022) found that O_3 tended to increase for a finer resolution simulation (~14 km vs. ~111 km over the CONUS) during the summer over urban areas using the Community Earth System Model (CESM)/Community Atmosphere Model with full chemistry (CAM-chem) model which was attributed to improvements in the spatial resolution of NO_x emissions resulting in less artificial dilution of NO_x and enhanced O_3 production. Similarly, Lin et al. (2024) found that a variable resolution global model (AM4VR with horizontal resolution of 13 km over the CONUS) had increased O_3 over urban areas compared to a fixed resolution model (AM4.1 with horizontal resolution of ~100 km globally). In particular for the Los Angeles Basin and Central Valley regions of California, Lin et al. (2024) found that the increased resolution of AM4VR led to better simulation of observed O_3 levels in these areas due the finer resolution model's ability to represent sharp spatial gradients in areas with NO_x-limited vs. NO_x-saturated O_3 production regimes. Given these previous results finding increased O_3 with finer resolution simulations, our results here finding higher biases in USA O_3 in the eastern US with coarser resolution should be taken to apply specifically to the CMAQ model results described here rather than as a general finding on the impact of model resolution on O_3 production."

Figure 6: the authors should present results for different seasons, not annual averages.

Our intention is that Figures 6 and 8 convey information about the spatial variability of the inferred model biases while Figures 7 and 9 convey information about the seasonal variability. We have, however, added new figures to the SI showing spatial maps of the seasonal average inferred model biases (Figures S5-S10 in the revised SI) so that these details are available.

Figure 13: What is the horizontal resolution of PA and EQUATES simulations presented in this figure? Are the differences driven by differences in model configurations or model resolution? The authors should show the comparison from the same configuration but at different resolutions.

The simulations referred to in Figure 13 are at 12 km resolution. This has been added to the figure and caption. The purpose of Figure 13 is to add context to the rest of Section 3.4 which deals exclusively with results from the 12 km simulations. Figure 13 is intended to provide additional context about how representative the results shown in Figures 11 and 12 for O_3 > 70 ppb are for the western US and eastern US regions more broadly. While similar results for other model resolutions are of general interest, this would not fit in with the overall theme of Section 3.4 which otherwise deals only with 12 km results. For these reasons, we have added the suggested figure to the SI (Figure S11 in the revised SI) to show the same results as in Figure 13 for 36 km model resolution (PA simulation only) and 108 km model resolution (both PA and EQUATES simulations).

3. Discussion on stratospheric contribution and CMAQ low-O3 bias in spring should be placed in the broader published literature, including those using dynamic stratospheric ozone tracers with explicit stratospheric chemistry and evaluation with intensive ozone profiling during western US field campaigns. The stratospheric contribution estimated by CMAQ appears to be much lower than the estimates from these prior studies:

A.O. Langford, R.J. Alvarez II, J. Brioude, R. Fine, M. Gustin, J.S. Holloway, M.Y. Lin, R.D. Marchbanks, R.B. Pierce, S.P. Sandberg, C.J. Senff, A.M. Weickmann, E.J. Williams, *Entrainment of stratospheric air and Asian pollution by the convective boundary layer in the Southwestern U.S.*, J. Geophys. Res., 122 (2), doi:10.1002/2016JD025987, 2017.

Lin M., A. M. Fiore , O. R. Cooper , L. W. Horowitz , A. O. Langford , Hiram Levy II , B. J. Johnson , V. Naik , S. J. Oltmans , C. Senff (2012): Springtime high surface ozone events over the western United States: Quantifying the role of stratospheric intrusions, *Journal of Geophysical Research*, 117, D00V22, doi:10.1029/2012JD018151

Langford, A.O., C.J. Senff, R.J. Alvarez II, J. Brioude, O.R. Cooper, J.S. Holloway, M.Y. Lin, R.D. Marchbanks, R.B. Pierce, S.P. Sandberg, A.M. Weickmann , E.J. Williams (2015): An overview of the 2013 Las Vegas Ozone Study (LVOS): Impact of stratospheric intrusions and long-range transport on surface air quality. Atmos. Environ, doi:10.1016/j.atmosenv.2014.08.040

Langford, A. O., Senff, C. J., Alvarez II, R. J., Aikin, K. C., Baidar, S., Bonin, T. A., Brewer, W. A., Brioude, J., Brown, S. S., Burley, J. D., Caputi, D. J., Conley, S. A., Cullis, P. D., Decker, Z. C. J., Evan, S., Kirgis, G., Lin, M., Pagowski, M., Peischl, J., Petropavlovskikh, I., Pierce, R. B., Ryerson, T. B., Sandberg, S. P., Sterling, C. W., Weickmann, A. W., and Zhang, L.: *The Fires, Asian, and Stratospheric Transport-Las Vegas Ozone Study (FAST-LVOS)*, Atmos. Chem. Phys., https://doi.org/10.5194/acp-2021-690, 2022.

Below is additional discussion of the stratospheric O_3 contribution in the EQUATES CMAQ simulations that has been added in Section 3.3. In these additions, the comparisons to previous work are mostly for modeling studies that have reported seasonal mean stratospheric contributions rather than work from field intensives that tend to focus on specific stratospheric intrusion events that are not directly comparable to the seasonal estimates.

"In the 12 km EQUATES simulations, the STRAT O_3 tracer averages 14 ppb in the western US during spring, with a maximum spring average across all western US grid cells of 17 ppb. Using the bias correction approach developed here, we find that the spring average STRAT O_3 in the western US is biased low by 3.5 ppb, resulting in an adjusted (i.e., bias corrected) estimate of western US spring average STRAT O_3 of 17 ppb. Consistent with the low bias in stratospheric O_3 suggested here, other CTMs have estimated higher stratospheric O_3 contributions compared to those simulated here with CMAQ. The spring average of stratospheric O_3 contributions estimated with the AM3 model has been estimated at 20-25 ppb (Lin et al., 2012a; Langford et al., 2015; Lin et al., 2015). The AM3 estimates of stratospheric O_3 have sometimes been estimated to be biased high (Lin et al., 2012a) and have also been shown to lead to overestimated springtime O_3 concentrations when used as boundary conditions for regional-scale CMAQ simulations (Hogrefe et al., 2018) but at other times have been estimated to be relatively unbiased based on evaluation against observations from intensive field studies (Langford et al., 2015). The stratospheric O_3 contribution simulated by AM3 has been previously found to be higher than that of the GEOS-Chem global model (Fiore et al., 2014). Using GEOS-Chem, Zhang et al. (2014) found the spring mean stratospheric O_3 influence in the Intermountain West to range from 8-10 ppb as estimated using the standard GEOS-Chem definition of stratospheric O_3 as described in Zhang et al. (2011) and, alternatively, found a spring mean of 12-18 ppb using a definition of stratospheric O_3 adopted from Lin et al. (2012a) (the same method used for the AM3 estimates reported here). Itahashi et al. (2020) previously found that the stratospheric O_3 representation in CMAQ was biased low in the free troposphere and suggested that improvements were needed to the CMAQ representation of stratosphere to troposphere transport. Our bias adjusted estimate of western US spring mean stratospheric O_3 (17 ppb) falls in between the estimates from the default GEOS-Chem representation (8-10 ppb) and from AM3 (20-25 ppb). As these are seasonal averages, the values are more representative of the continual entrainment of stratospheric air into the troposphere rather than episodic deep stratospheric intrusion events."

Other comments:

Lines 29-30: "USB O3 … is a larger portion of total observed O3 as anthropogenic precursor emissions decline". This statement needs a few references, such as Lin et al. (2017):

Lin, M.., W. Horowitz, R. Payton, A.M. Fiore, G. Tonnesen (2017). *US surface ozone trends and extremes from 1980 to 2014: Quantifying the roles of rising Asian emissions, domestic controls, wildfires, and climate*. Atmos. Chem. Phys., doi:10.5194/acp-17-2943-2017

We have added citations for this sentence, including the one suggested by the reviewer.

Lines 38-48: Need references. Could also discuss the difficulty to separate the anthropogenic and natural driver of wildfire impacts on ozone air quality, as ozone production is enhanced due to mixing of wildfire VOC emissions with urban NOx?

Several references have been added throughout this paragraph. The point about wildfire impacts that the reviewer raises is also relevant to this passage, so the following has been added to this section:

"Wildfires are treated as USB O_3 sources, but the impacts of wildfires on O_3 can be affected by US anthropogenic emissions when VOCs from fires are transported over NO_x-rich urban areas, leading to enhanced O_3 production (Jaffe et al., 2013; Langford et al., 2023; Rickly et al., 2023)."

Table 1: The referee agrees with Referee #1 that the authors should list more detailed information regarding model version, simulations types, horizontal and vertical resolution, US anthropogenic emissions, international emissions, fire emissions (including temporal frequency and injection height), and other natural emissions.

Many of these details are available from Tables S4 and S5. Information on the vertical layer structure and more details on the modeling domains have now been added to Table S5. We have also added additional details given below in Section 2.1 of the manuscript. (Note this response is identical to the response to reviewer #1 comment marked "2. Methods")

Vertical structure:

"Both the PA and EQUATES simulations use a 44-layer vertical structure for hemispheric scale applications (at 108 km resolution) and a 35-layer vertical structure for continental (i.e., 36 km and 12 km resolution) applications with a vertical extent from the surface to 50 hPa and a surface layer height of approximately 20 m for both the hemispheric and continental configurations (see Mathur et al. (2017) for more details on these vertical layer structures)."

Model configuration (chemical mechanism and meteorological model version):

"Besides the addition of halogen chemistry, there are other differences in the chemical mechanisms used for each set of simulations. The mechanisms used for the hemispheric simulations were cb6r3_ae6_aq for the PA simulations and cb6r3m_ae7_kmtbr for the EQUATES simulations. The part of the mechanism name labeled cb6r3m indicates additional chemistry relevant in marine environments (the halogen chemistry described above); ae6 and ae7 indicate the version number for chemistry relevant to aerosols; aq and kmtbr indicate different treatments of cloud chemistry. The chemical mechanisms used for continental-scale PA and EQUATES simulations (cb6r3_ae6nvPOA_aq and cb6r3_ae7_aq) also differ in their representation of organic aerosols (Murphy et al., 2017; Pye et al., 2019; Qin et al., 2021; Appel et al., 2021) which could affect O_3 concentrations. Different versions of WRF (v3.8 for PA simulations and v4.1.1 for EQUATES simulations) employed may also contribute to differences in O_3."

Emissions and stratospheric O_3:

"Emission inputs also differ between the PA and EQUATES simulations. Different US anthropogenic emission inventories were used for the simulations. The PA simulations used an early version (sometimes called the "alpha" version) of a 2016 emissions modeling platform developed by the National Emissions Inventory Collaborative (US EPA, 2019a). The EQUATES simulations used an inventory that was developed as part of the broader EQUATES framework to model a long timeseries using consistent methods for emissions estimates (Foley et al., 2023). For emissions in Canada and Mexico, both sets of simulations use emission inventories developed by the respective national governments, though the EQUATES simulations use more recent inventories (as described by Foley et al. (2020)) than the PA simulations (as described by US EPA (2019a)). Both the PA and EQUATES simulations use the Tsinghua University inventory of emissions in China (Zhao et al., 2018). For other countries, both sets of simulations use the Hemispheric Transport of Air Pollution (HTAP) v2.2 inventory (Janssens-Maenhout et al., 2015) with scaling factors derived from the Community Emissions Data System (CEDS) (Hoesly et al., 2018) to account for yearly changes. Differences in the anthropogenic emissions used in the two model configurations are expected to contribute to differences in simulated O_3, most notably for the different US anthropogenic emissions since we focus here on O_3 in the US.

For hemispheric-scale simulations, biogenic VOC emissions are from the Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1) (Guenther et al., 2012). The PA simulations additionally replace MEGAN emissions with emissions from the Biogenic Emission Inventory System (BEIS) (Bash et al., 2016) over North America (USEPA, 2019b). The EQUATES MEGAN emissions are obtained from a compilation by Sindelarova et al. (2014). Soil NO_x emissions for the PA hemispheric simulations are also from MEGAN with replacement by BEIS soil NO_x over North America. Soil NOx emissions for the hemispheric EQUATES simulations are from a dataset by the Copernicus Atmosphere Monitoring Service (CAMS, 2018) based on methods by Yienger and Levy (1995). Lightning NO emissions for both the PA and EQUATES hemispheric simulations are from monthly climatology obtained from the Global Emissions Initiative (GEIA) and are based on Price et al. (1997). Lightning NO_x was not included in the PA continental-scale simulations, while lightning NO_x for the EQUATES continental-scale simulations is calculated using an inline module in CMAQ (Kang et al., 2019). For both PA and EQUATES, wildfire emissions outside of North America are based on the Fire Inventory from NCAR (FINN) v1.5 (Wiedinmyer et al., 2011) which provides day-specific fire emissions. Wildfires are vertically allocated with 25% of emissions distributed to the lowest two layers (~0-45 m), 35% distributed to layers 3-9 (~45-350 m), and the remaining 40% distributed to layers 10-19 (~350-2000 m) as described in the Technical Support Document for northern hemispheric emissions (US EPA, 2019b). Wildfire emissions within North America are based on the Hazard Mapping System (HMS) fire product which provides day-specific fire activity data. Emission processing for North American wildfires is further described in the Technical Support Document for North American emissions (US EPA, 2019a) (applicable to PA simulations) and Foley et al. (2023) (applicable to EQUATES simulations). Although the methods are similar, North American wildfire emissions may differ between PA and EQUATES based on the specific fire activity data that was used in each case. Fire plume injection height for North American fires is determined by an inline plume rise algorithm in CMAQ based on fire heat content (see e.g., Wilkins et al. (2022) for more details on fire plume injection height in CMAQ). Stratospheric O_3 in both the PA and EQUATES simulations is from the PV parameterization by Xing et al. (2016) (described in more detail above) in the hemispheric simulations. Stratospheric O_3 in the continental-scale simulations only comes from any stratospheric O_3 inherited from the lateral boundary conditions provided by the hemispheric simulations."

For all of the figures, please indicate in the figure caption whether MDA8 O3 or 24-h O3 is shown. There are some discussions of the metric in the first paragraph of Section 2.1. But it is much easier for readers if you label them as "MDA8 O3" directly in the figure captions.

Captions have been updated to indicate MDA8 O_3.

Figures 2 to 5: Results in the maps look pretty similar in their current form. Please use a different colorbar so that the spatial distribution of different model configurations can be better illustrated!

Main text and SI figures have been updated to use different color scales.