We thank the reviewer for the thoughtful and detailed comments. We have revised this manuscript carefully based on the comments. In particular, Fig. 2, Fig. 4, Fig. 7 and Fig 9 were redrawn to facilitate the readers with color vision deficiencies as suggested by the editorial team. Fig.4F1-F5 and Fig. 9F1-F5 were adjusted to better match the description in the captions and Fig. S6 was adjusted by using log scales as suggested by the reviewer.

Below we respond to the individual comments.

Reviewer #1

Question: The comments of the reviewers have been addressed, and the paper has been improved so that it is clearer and more readable. The analysis is substantially more thorough than in the original manuscript, and the results are now more robust and more useful. However, there are still some weaknesses that need to be addressed before the paper is appropriate for publication.

Answer: Thank the reviewer again for the constructive comments! The manuscript has been revised based on the comments.

Question: Line 57: "combination of multiple species": please specify the species included in Ox, to help readers who are not familiar with GEOS-Chem. (This information is provided on line 200, but would be better at this earlier point).

Answer: The definition of O_x was moved forward to the Introduction Section.

Question: Line 185: How are the archived PO3 and LO3 terms applied to the single tracer? Is the LO3 term kept as a first-order loss term (LO3/[O3]) so that the original [O3] does not need to be archived? A sentence or two on the methods is needed for publication in GMD.

Answer: As indicated by the reviewer, the LO_3 that was read in the single O_3 tracer simulation is $LO_3/[O_3]$. The method has been clarified in the revised version (Lines 171-173):

"The GEOS-Chem full chemistry simulations with the updated KPP module were then performed to produce PO_3 (unit kg cm^{-3} s^{-1}) and relative LO_3 (i.e., $LO_3/[O_3]$ with unit cm^{-3} s^{-1}) every 20 minutes".

Question: Tagged-Ox exceeds Tagged-O3, but this is because the additional NOx/NOy species are included. The concentrations will be a lot closer if these are removed! The comparison doesn't indicate that the Tagged-O3 run is better, merely that it is more suitable for direct comparison with observed O3.

Answer: The major difference between the tagged-O_3 and tagged-Ox is emphasized in the revised version:

"it may not be an ideal choice to perform O_3 simulations based on the tagged-O_x mode because O_x is the combination of multiple species (O_x=O_3+NO_2+2NO_3+3N_2O_5+HNO_3+HNO_4+peroxyacylnitrates) and thus cannot be accurately compared with O_3 observations" in the Introduction Section.

"In contrast, the O_x concentrations provided by the tagged-O_x mode are higher than the O_3 concentrations by approximately 6 ppb, and the relative difference can reach 40% in the winter, which is thus not suitable for direct comparison with observed O_3" in the Conclusion Section.

Question: It is not surprising that the Tagged-O3 run can match the full model run, as PO3 and LO3 are derived directly from this run. A better test of the tagged mode would need to explore what happens under ozone changes (e.g., those associated with assimilation), as with increasing ozone changes the approach will lose accuracy. The approach described is good for representing near-current conditions, but is not suitable under substantially different conditions; the paper should be clear about this limitation.

Answer: Thank the reviewer for pointing out this issue! The limitation of the tagged-O_3 simulation was elaborated in the Conclusion Section: "Despite these advantages, it should be noted that the linear chemistry assumption by reading the archived PO_3 and LO_3 implies single O_3 tracer mode is good for representing near-current O_3 chemical conditions, particularly, for scientific issues associated with the sources and transport of tropospheric O_3 as well as assimilations in this work and the companion paper (Zhu et al., 2023). More cautious applications are suggested under substantially different O_3 chemical conditions as the linear chemistry assumption could not be satisfied".

Question: Fig 4 caption: note that F1-F5 show the effect of *removing* O3 formation over the NCP; the "effects of O3 formation" would be given by showing #1-#4. Note that a less extreme color scale (e.g., 40 to -40) would show up the effects outside the NCP more clearly.

Answer: Fig. 4F1-F5 and Fig. 9F1-F5 have been redrawn by showing experiments #1-#4. The color scale was adjusted to 40 to -40.

Question: Fig 5: 5 regions are defined in Fig 1, but 6 regions are shown here; it would be helpful to remind the reader that "E. China" refers to the whole domain of interest here.

Answer: Thank the reviewer for this suggestion! The domain definition of E. China was reminded in the captions of Table 2 and Fig. 5 in the revised version.

Question: The final section is a little short on analysis; results are described, but the consequences and implications of them are not identified or explored in much detail.

Answer: The discussion in the Conclusion Section has been adjusted to emphasize more on the consequences and implications of the analysis.

Question: Note that interannual variability should not affect long-term trends, it just affects assessment of trends over very short time periods such as the 6 years considered here.

Answer: We agree with the reviewer that it could be more accurate to use the term "interannual trends in background O_3". However, the interannual trend of background O_3 is not the target of this work and is not evaluated in our analysis. Consequently, we use the term "interannual variability in background O_3" because we found it may not be robust enough to make a conclusion about the interannual trends of background O_3.

Supplement

Question: Uncertainty analysis: "...randomly drawing N data points from the full set of N data points...". Should this be "with replacement"? If so, there is duplication. If not, the sampling just uses all points. Some clarification on the method used is needed here.

Answer: The bootstrapping method allows drawing individual data points multiple times to represent the error due to random sampling. It has been clarified in the revised description: “individual data points may be drawn multiple times”.

Question: Fig S6: given the mass units of PO3, it would be better to present these graphs on a log-log scale (i.e., use a log scale on the X-axis, too).

Answer: Thank the reviewer for this suggestion! Fig. S6 has been adjusted by using log scales.