Reply to Comments by Anonymous Referee #1 on:

Assessing the relative impacts of satellite ozone and its precursor observations to improve global tropospheric ozone analysis using multiple chemical reanalysis systems

Sekiya, T., Emili, E., Miyazaki, K., Inness, A., Qu, Z., Pierce, R. B., Jones, D., Worden, H., Cheng, W. Y. Y., Huijnen, V., and Koren, G.
EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-2426
Discussion started: 20 August 2024; Discussion closes 01 Oct., 2024

We would like to thank anonymous referee #1 for their careful reading of our manuscript and for providing valuable comments, which have greatly improved the manuscript. We have revised the manuscript accordingly, and the main changes are summarized below:

1) An analysis was added to highlight the seasonal cycles in the improvements in ozone analysis by data assimilation (Figure 5 in the revised manuscript).
2) A more in-depth discussion was added on the impacts of discontinuity in the current satellite observing system on tropospheric ozone analysis, as well as potential strategies to address this issue after the termination.
3) An analysis was added to the Supplementary material to explain the reasons behind the large differences between TCR-2 and GEOS-Chem adjoint based on only the OMI NO2 assimilation.

Individual comments (in black) and point-by-point replies (in blue) are listed below. Revised text (*italicized font*) from the updated manuscript are in quotes below.

Note that we have added line types in Figure 4 in the revised manuscript to assist readers with color vision deficiencies according to the comments from Editorial Support.

The manuscript assesses the improvement in tropospheric ozone achieved through the assimilation of satellite observations of ozone and some of its precursors (NO2 and CO) in five reanalyses products compared to ozonesonde observations and satellite-based tropospheric ozone estimates. The study also distinguishes between the effects of directly assimilating ozone observations and assimilating only precursor observations. The main conclusion is that assimilation improved the consistency among the reanalyses products, though the impact varies among systems because of differences in the forecast model and assimilation configurations.

The study makes a significant contribution to the field by demonstrating both the capabilities and limitations of the current observing system in constraining the distribution of tropospheric ozone. By examining multiple reanalysis products, it goes beyond most previous studies that have often considered just a single product. The results are presented in a balanced manner, figures and tables are used where necessary, and the text is well-written.

We appreciate your careful review and positive comments.

The following are a few aspects that could be discussed further to help improve the manuscript:
i) It would be helpful to clarify why none of the recently developed satellite-based tropospheric column products (eg. IASI-GOME2 product) were used for evaluation. While these products aren't available for 2010 (the study's evaluation period), the reanalysis datasets extend to more recent years.

The analysis in the present study primarily focuses on 2010. Among the satellite-based tropospheric ozone column (TOC) products, the OMI-MLS TOC product was used because it covers the period (i.e., 2010) when all the reanalysis datasets and control runs, including the experimental reanalysis data, were available.
We have modified the relevant sentence as follows:
(p. 8, lines 222–225)
"*..., we used global distributions of the tropospheric ozone column (TOC) derived from OMI-MLS (Ziemke et al., 2006), which is ... and covers the period when all the reanalysis datasets and control simulations were available.*"

Additionally, three of the five reanalysis datasets were available in 2016. Thus, we compared the available reanalysis datasets with recently developed satellite products (Fig. 1). We used the retrievals of lowermost partial ozone column (0–3 km altitude) derived from the IASI+GOME-2 product (Cuesta et al., 2013), which have enhanced sensitivity in the lowermost troposphere because of the multi-spectral approach using the IASI and GOME-2 measurements in the thermal infrared and ultraviolet spectral domain. The ozone analysis fields were used without application of averaging kernels, as done by Okamoto et al., 2023). Compared to IASI+GOME-2, ozone analyses obtained from

CAMSRA, TCR-2, and GEOS-Chem commonly showed negative biases typically within 3–6 DU. Spatial correlation coefficients between IASI+GOME-2 and reanalysis datasets ranged within 0.53–0.58. The comparisons show commonly systematic bias patterns of all the reanalyses against IASI+GOME-2 products, which were inconsistent with the comparisons against ozonesonde observations. These results suggest a potential issue with the IASI+GOME-2 data or validation approach (e.g., not applying averaging kernels) and indicate that it does not provide helpful validation for this study. While recently developed satellite products from instruments, such as TROPOMI, CrIS, and AIRS-OMI are still under validation, it is meaningful to note the value of incorporating these datasets into future validation studies.

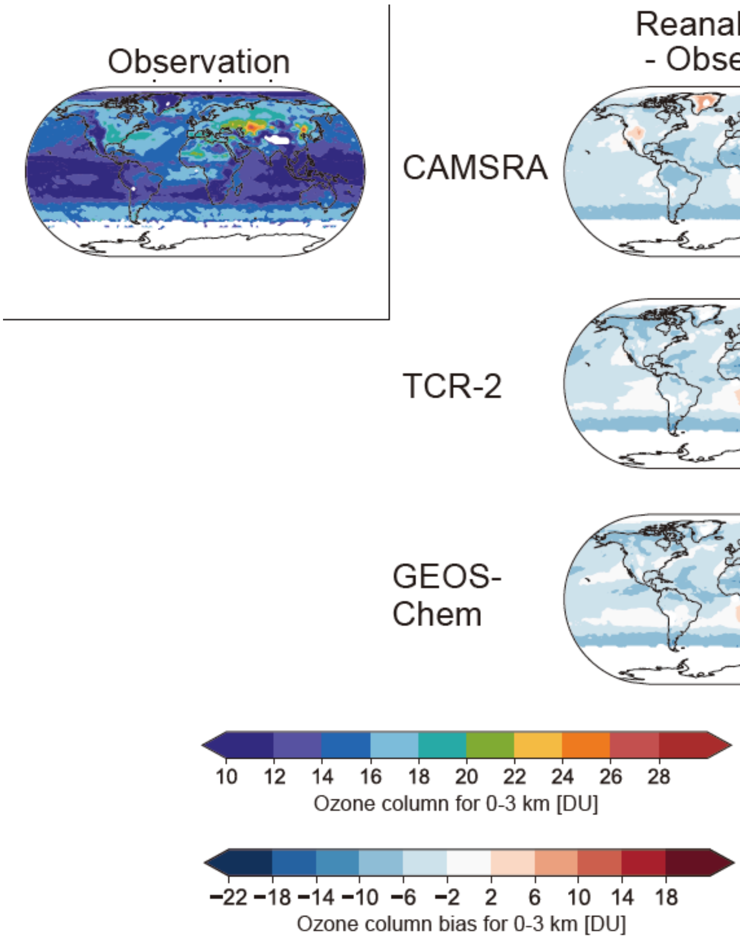

Figure 1. Global distributions of partial ozone column for 0–3 km altitude derived from the IASI+GOME2- observations (first column) and mean bias of the CAMSRA (first row), TCR-2 (second row), and the GEOS-Chem-adjoint (third row) relative to IASI+GOME-2 during June, July, and August 2016. The unit is the Dobson unit (DU).

ii) The analysis of the comparison of the reanalysis products with the OMI-MLS tropospheric column data could be expanded. In particular, it is interesting that assimilation introduces a bias in the models that was absent in the control runs without data assimilation (Figure 3). Also, the high biases seen in the comparison with OMI-MLS are not seen in the comparison with ozonesonde measurements.

We have added a comparison between OMI-MLS and ozonesonde data to Figure 3. This comparison indicates that OMI-MLS presented smaller TOC than ozonesonde observations for many cases, and these findings are consistent with the positive biases observed for the reanalysis relative to OMI-MLS. The following sentence has been added in the revised manuscript:
(p. 10, lines 287–289)
"*After data assimilation, all chemical reanalysis products consistently revealed positive biases of 5–10 DU relative to OMI-MLS from the tropics to the mid-latitudes of both hemispheres. However, this result is inconsistent with the reanalysis comparison results against ozonesonde measurements (see Section 3.2.2). Considering ozonesonde measurements as ground truth, part of the positive bias relative to OMI-MLS can be attributed to smaller TOC in the OMI-MLS data (by 3.7 DU on average) compared to ozonesonde observations, as confirmed in Figure 3 and reported by Gaudel et al. (2024).*"

iii) It would be useful to discuss how the loss of instruments like MLS, OMI, and MOPITT upon decommissioning of Aura and Terra might affect the quality of tropospheric ozone reanalysis datasets in the future.

We have added the discussion on the continuity of constraints on tropospheric ozone by satellite observations after the decommission of current satellites as follows:
(p. 15, line 472–p. 16, line 480)
"*The current satellite observing system provides unique and essential information that are essential for improving ozone analysis throughout the troposphere and stratosphere, as demonstrated in this study. However, the termination or retirement of current instruments such as OMI, MLS, MOPITT, and IASI in the near future may impact the ability to constrain whole troposphere ozone profiles effectively. More recent and future satellite measurements, such as TROPOMI, CrIS, and IASI-New Generation (NG) offer*

the potential to maintain or even improve constraints on tropospheric ozone and its precursors. For instance, advances in TROPOMI NO_2 assimilation compared to OMI NO_2 assimilation, as demonstrated by Sekiya et al. (2022), highlight these capabilities. For UTLS ozone analysis, several measurements, such as OMPS and SAGE III, continue to provide valuable profile measurements."

iv) It would also be useful to discuss whether new types of observations (e.g. profiles in the UTLS) are needed to better constrain tropospheric ozone. Would continuing the current observing system be sufficient, provided that the information from these observations is fully utilized?

The discussion what information will be needed to improve ozone analysis has been added, especially for UTLS ozone profiles, as follows:
(p. 16, lines 480–485)
"*Nevertheless, the uniqueness of MLS to observe through clouds and aerosols and a wide range of trace gases remains powerful in constraining tropospheric ozone profiles and chemistry system. The development of follow-on missions, such as the Atmospheric Limb Tracker for Investigation of the Upcoming Stratosphere (ALTIUS) (Fussen et al., 2019) and the Stratosphere Troposphere Response using Infrared Vertically-Resolved Light Explorer (STRIVE), to fill the gaps is highly desirable for maintaining high-quality tropospheric ozone analysis. Assessing the observational impacts of these instruments through data assimilation is expected to provide critical information for optimizing the observation system.*"

References

Cuesta, J., Eremenko, M., Liu, X., Dufour, G., Cai, Z., Höpfner, M., von Clarmann, T., Sellitto, P., Foret, G., Gaubert, B., Beekmann, M., Orphal, J., Chance, K., Spurr, R., and Flaud, J.-M.: Satellite observation of lowermost tropospheric ozone by multispectral synergism of IASI thermal infrared and GOME-2 ultraviolet measurements over Europe, Atmos. Chem. Phys., 13, 9675–9693, https://doi.org/10.5194/acp-13-9675-2013, 2013.

Okamoto, S., Cuesta, J., Beekmann, M., Dufour, G., Eremenko, M., Miyazaki, K., Boonne, C., Tanimoto, H., and Akimoto, H.: Impact of different sources of precursors on an ozone pollution outbreak over Europe analysed with IASI+GOME2 multispectral satellite observations and model simulations, Atmos. Chem. Phys., 23, 7399–7423, https://doi.org/10.5194/acp-23-7399-2023, 2023.

Reply to Comments by Anonymous Referee #2 on:
Assessing the relative impacts of satellite ozone and its precursor observations to improve global tropospheric ozone analysis using multiple chemical reanalysis systems
Sekiya, T., Emili, E., Miyazaki, K., Inness, A., Qu, Z., Pierce, R. B., Jones, D., Worden, H., Cheng, W. Y. Y., Huijnen, V., and Koren, G.
EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-2426
Discussion started: 20 August 2024; Discussion closes 01 Oct., 2024

We would like to thank referee #2 for their careful reading of our manuscript and valuable comments, which have greatly improved the quality of the manuscript. We have revised manuscript accordingly, and the main changes are as follows:

4) An analysis was added to highlight the seasonal cycles in the improvements in ozone analysis by data assimilation (Figure 5 in the revised manuscript).
5) A more in-depth discussion was added on the impacts of discontinuity in the current satellite observing system on tropospheric ozone analysis, as well as potential strategies to address this issue after the termination.
6) An analysis was added to the Supplementary material to explain the reasons behind the large differences between TCR-2 and GEOS-Chem adjoint based on only the OMI NO2 assimilation.

Individual comments (in black) and point-by-point replies (in blue) are listed below. Revised texts (*italicized font*) from the revised manuscript are in quotes below.

Note that we have added line types in Figure 4 in the revised manuscript to assist readers with color vision deficiencies according to the comments from Editorial Support.

In "Assessing the relative impacts of satellite ozone and its precursor observations to improve global tropospheric ozone analysis using multiple chemical reanalysis systems," Sekiya and coauthors compare 5 different global chemical reanalysis systems and the ozone in those systems against ground, ozonesonde, and space based observations. They evaluate assimilating total column ozone, ozone profiles, tropospheric ozone, and precursors of ozone, including total column CO, profiles of CO, and tropospheric NO_2 columns. For one system, TCR-2, authors test the impacts of different combinations of assimilated observations on the reanalysis. Authors conclude that UTLS ozone was most improved by assimilating total column and tropospheric ozone, while middle troposphere ozone was most improved by assimilating precursor observations. Lower troposphere ozone was improved by assimilating precursor observations. Across all reanalysis systems, the influence of assimilation on surface ozone was small, and in some cases made biases worse. Authors note that comparing across systems with different settings and assimilated species introduces uncertainties.

Overall, results will be of interest to the readers of ACP and the conclusions authors draw from their analysis will be informative for future reanalyses. The consistent evaluation against observations is a useful contribution. Some of the conclusions are not very clear, making it difficult to find the primary contributions of the work. The different settings in the reanalysis systems makes it difficult to draw specific conclusions in some cases, but this is probably unavoidable.

We appreciate your careful review. In response to the above comment indicating that some of the conclusions were not very clear, we have revised the last paragraph of the Conclusion section to clarify our intended meaning:
(p. 17, line 528–p. 18, line 544)
"*In conclusion, the simultaneous assimilation of satellite measurements of stratospheric and tropospheric ozone, and its precursors has proven to be an effective approach for improving the entire tropospheric ozone analysis. Despite variations in forecast model performance, assimilated observations, and data assimilation settings across the reanalysis systems used in the intercomparison, data assimilation greatly improved consistency among reanalysis products as well as with independent observations. The overall good agreement suggests that the highly accurate reanalysis datasets are valuable for advancing our understanding of atmospheric composition variations and can also inform discussions on the development of observing systems. Meanwhile, multi-system OSEs qualitatively demonstrated that ozone analyses from the middle troposphere to the lower stratosphere are improved through the assimilation of satellite ozone observations, while analyses from the lower to middle troposphere benefit from the assimilation of ozone precursors. However, the impacts of ozone assimilation from satellite observations on ozone analysis varies widely across the reanalysis systems, suggesting that individual results may introduce biases when evaluating the value of specific observing systems. This underscores the importance of employing multiple systems to ensure robustness in assessing individual observing systems. Furthermore, these findings highlight the need to*

account for the forecast model performance and data assimilation configurations when conducting OSEs/OSSEs. Such considerations, which have been lacking in previous observing system impact assessments, are essential to provide unbiased insights for designing future observing systems. Meanwhile, to draw more robust conclusions about the system dependence of observing system impact assessments, further studies involving intercomparisons of OSEs conducted with a more consistent and improved protocol compared to the present study are needed."

General comments

The structure of section 2.1 and its subsections makes it difficult to follow and therefore difficult to understand the differences between the reanalysis systems. Could authors adjust this section so that the same key words are used across the sub sections for key components? For example, it would be helpful to clearly mention the forward model, a priori emissions, and simulations settings used in each system using the same key words.

We have modified the descriptions of individual reanalysis datasets from Section 2.1.1 to 2.1.5. The structures were not completely consistent because of the differences between reanalysis systems. However, we used common keywords and applied the same order of description whenever possible in the revised manuscript.

For example, we provided general overview of each reanalysis in the first paragraph of individual sections, as follow:

(p. 5, lines 125−128)

"*Tropospheric Chemistry Reanalysis version 2 (TCR-2) provides both emissions and atmospheric abundance of various chemical species from the assimilation of multi-constituent measurements from multiple satellite instruments during 2005–2021 (Miyazaki et al., 2020a). This reanalysis products were developed under the Multi-mOdel Multi-cOnstituent Chemical data assimilation (MOMO-Chem) framework (Miyazaki et al., 2020b).*"

Also, we arranged the description of individual reanalysis datasets as follows:

(1) The assimilated satellite observations were described in the second paragraph.
(2) The descriptions of data assimilation systems (e.g., horizontal resolution, data assimilation technique) were provided in third paragraph.
(3) The forecast models' description in the fourth paragraph was followed by the specification of a priori emission inventories (if used).

In the discussion, authors say that “Furthermore, the present study shows that the spread of data assimilation impacts among multiple systems can be used to evaluate whether the observing system impacts are dependent on the reanalysis system,” and that “These findings should lead to more robust assessments of the observing system impacts.” It would be helpful to clearly summarize and/or enumerate what those findings are,

We have revised the sentences to clarify the meaning of “these findings” as follows:
(p. 14, line 435–p. 15, line 440)
“*The results obtained from multi-system OSEs have two important implications for the development of a future satellite constellations as follows: (1) Integrating measurements of ozone and its precursors is an effective way to improve the entire profile of ozone in the troposphere and lower stratosphere, as consistently suggested by previous single-system OSE studies (e.g., Miyazaki et al., 2019). This finding highlights the great value of the current satellite constellation. (2) The spread obtained for data assimilation impacts across multiple systems, including notable differences in certain areas, provide key insights for determining the influence of reanalysis system choices on the observing system impact assessments.*”

and specifically how the current study shows whether the observing system impacts are dependent on the reanalysis system.

We have added a more in-depth discussion on how DA increments in ozone depend on different systems. Also, remaining biases in the reanalyses have been described as follows:
(p. 15, lines 440–445)
“*The absolute values of DA increments largely varied, ranging 0–21%, 0–22%, and 0.1–19% in the UTLS, middle, and lower troposphere, respectively. However, the remaining biases of reanalyses against ozonesonde observations were within ±2.5%, ±3.5%, and ±7.0% in the UTLS, middle, and lower troposphere, respectively, except for GEOS-Chem adjoint. These differences could be related to a variety of factors, including different retrieval algorithms, data assimilation settings, which are applied to fully exploit observational information (e.g., assumption of background error covariance), and different model performance.*”

Ozone exhibits seasonal biases in many modeling systems, but here authors present annual mean results only. It would be ideal to present seasonal results to understand the impact of different assimilated observations in different seasons, but that may be out of the scope of this analysis. At a minimum, authors should mention this.

We agree with the reviewer that seasonality is an important component of model biases. Similar comments were also raised by other reviewers, prompting us to include a more detailed discussion. The seasonal dependence of model biases and their improvements by data assimilation (reanalysis) has been added to Figure 5, along with corresponding descriptions in section 3.2.2:

(p. 10, line 305–p. 11, line 308)
"*As shown in Figures 5 and S1, over the northern mid and high latitudes, RAQMS showed larger seasonal amplitudes in model bias, with a maximum in boreal spring, compared with CAMS, TCR-2, and IASI-r. Over the southern mid and high latitudes, CAMS and RAQMS showed larger negative biases in austral summer and fall compared to other reanalyses, while TCR-2 and IASI-r exhibited maximum positive biases in austral spring.*"

(p. 11, lines 312–314)
"*The seasonal variations of model biases were approximately 20% larger in IASI-r compared to the other models. IASI-r showed maximum negative biases over the northern midlatitudes and maximum positive biases over the southern high latitudes during summer.*"

(p. 11, lines 318–320)
"*IASI-r exhibited larger biases compared to other models over the mid and high latitudes in summer in both hemispheres, whereas RAQMS and TCR-2 showed larger negative biases over the mid and high latitudes in winter in both hemispheres.*"

(p. 11, line 325)
"*Seasonally dependent biases were also improved by data assimilation over all the latitude bands.*"

(p. 11, lines 331–333)
"*The seasonal variations in model biases were also reduced. However, multi-system mean*

biases in winter remained over the high latitudes in both hemispheres, likely due to the limited number of assimilated observations in these regions during winter."

(p. 11, line 339)
"*In these regions, the seasonal dependency of model biases was not improved by data assimilation.*"

In all figures besides 6 and 7, the fonts in the colorbars and legends are too small and are unreadable. Please make the fonts larger.

For Figures 1–3 and 5–8 we enlarged the individual panels, and removed the labels from each panel, showing only one label for each column and row. Also, the color bars on each panel were also removed and shown once below the panels.

In the ozonesonde plots, the lines are too small to distinguish colors and line types. Please make the lines larger and more easy to distinguish.

For Figures 4 and 9, we also enlarged the panels and made the lines thicker.

Specific comments
Line 4-5: This sentence is hard to follow, please adjust for clarity.

This part was modified as follows:
(p. 1, lines 4–5)
"*Observing system experiments (OSEs) were conducted with multiple reanalysis systems under similar settings to evaluate the impacts of reanalysis system selection on the quantification of observing system values.*"

Line 10: What values are these percentages referring to?

We have added an explanation on the percentages as follows:
(p. 1, lines 10–11)

“*increases in global lower tropospheric ozone by 0.1% in GEOS-Chem and 7% in TCR-2, with only NO_2 assimilation*”

Lines 32-34: This is a very general statement that I’m not sure can be attributed only to the review paper cited here. Consider rewording.

This part has been modified as follows:
(p. 2, lines 32–35)
“*A recent study indicated the complex impacts of various precursors on tropospheric ozone variability at regional to global scales (Elshorbany et al., 2024). Satellite HCHO and NO2 retrievals were used to diagnose ozone chemical regimes and relate them to recent ozone increasing trends over China (Lee et al., 2022; Ren et al., 2022) and reversal ozone weekend effects in the US cities (Jin et al., 2020).*”

Line 226: How was this regridding done? It seems import enough to mention this earlier and briefly describe how it was done.

The interpolation using inverse distance weighting and bilinear interpolation were used for regridding. We added the methodology as follows:
(p. 9, lines 252–254)
“*For the comparison, ozone analysis fields in CAMSRA and TCR-2 were regridded from their original model grid points onto a 2° ×2° grid using inverse distance weighting, while those in GEOS-Chem and RAQMS were regridded using bilinear interpolation.*”

Section 3.1: It seems worth noting that GEOS-Chem control minus reanalysis is essentially zero, different from other systems,

We have referenced minor impacts on ozone analysis in GEOS-Chem as
(p. 9, lines 267–268)
“*Data assimilation in GEOS-Chem adjoint has a minimal impact on free tropospheric ozone compared to the other systems.*”

and describing why there is essentially no change in the GEOS-Chem reanalysis.

The reason for the small change has been added to the Discussion section and the Supplementary material.
(p. 15, lines 447–452)
"*These results indicate that the assessment of observing system impacts is sensitive to the choice of data assimilation system. This difference may reflect the different a priori surface NOx emissions between TCR-2 (46.48 Tg N yr*$^{-1}$*) and GEOS-Chem (52.20 Tg N yr*$^{-1}$*) (Table S1) as well as differing model resolutions (Sekiya et al., 2021). Additionally, systematic biases among the assimilated products derived from different retrieval algorithms may contribute to these differences (Qu et al., 2020). GEOS-Chem adjoint shows larger data assimilation impacts on ozone when assimilating the DOMINO product than the NASA standard product (Figure S2).*"

Line 267: "differences in procedures for computing TOC" – how different are the procedures between the models, and roughly how much difference do you expect this to contribute?

The processes of computing TOC between the models were mostly consistent. However, differences in vertical discretization between the observations (MLS resolution of 2–3 km) and model (less than 1 km) can affect the TOC computation, especially in cases where a sharp gradient of ozone occurs around the tropopause. Evaluating the magnitude of these errors is challenging due to the lack of true ozone profiles in many locations, and sch an evaluation was not performed in this study. The following statement has been added in the revised manuscript:
(p. 10, lines 298–300)
"*Vertical resolution of the compared data differed largely around the tropopause (i.e., MLS resolution of 2–3 km and model resolution of ≤ 1 km), which can affect the computation of TOC when a sharp ozone gradient occurs and may lead to discrepancy in the comparison (Schoeberl et al., 2007).*"

Line 268-271: The discussion about positive bias in or against OMI-MLS profiles is confusing and hard to follow, please consider rewording.

We have modified the discussion as follows:
(p. 10, l. 288–296).
"*After data assimilation, all chemical reanalysis products consistently revealed positive biases of 5–10 DU relative to OMI-MLS from the tropics to the mid-latitudes of both hemispheres. However, this result is inconsistent with the reanalysis comparison results against ozonesonde measurements (see Section 3.2.2). Considering ozonesonde measurements as ground truth, part of the positive bias relative to OMI-MLS can be attributed to smaller TOC in the OMI-MLS data (by 3.7 DU on average) compared to ozonesonde observations, as confirmed in Figure 3 and reported by Gaudel et al. (2024). Data assimilation improved the spatial correlation coefficients to > 0.83 for all systems, demonstrating the usefulness of the reanalysis products and the value of the OMI-MLS data for the evaluation of TOC spatial distributions. The remaining discrepancies underscore the challenges in improving tropospheric ozone analyses through the assimilation of precursors, stratospheric profiles, or column ozone measurements.*"

Lines 349-351: This sentence is difficult to follow, what are the numbers referring to?

We have modified this sentence as follows:
(p. 13, lines 391–396)
"*The surface ozone change was smaller in IASI-r than TCR-2, partly because the prescribed background error covariance in IASI-r (10%) was lower than the typical background error in TCR-2 (15–17%) in the lower troposphere. In TCR-2, the background error covariance was dynamically estimated through ensemble model simulations at each location and time, which is expected to provide more accurate estimates of background error. Other reasons for significant differences near the surface could be related to the altitude-dependent sensitivities of the assimilated observations (i.e., averaging kernels).*"

Line 413: Wouldn't chemical systems be stiff, and more likely to be numerically "unstable" (rather than "stable" as the authors write)? Perhaps I am misunderstanding, and authors can clarify in that case

We intended to mention that the chemical system is not chaotic as in meteorological dynamics. Thus, we have rephrased this term to "*non-chaotic.*" (p. 15, line 454)

Line 442-443: It is unclear what exactly is being reduced and by how much, please adjust for clarity

This sentence has been modified as follows:
(p. 16, lines 505−507)
"*After data assimilation, the RMSEs against global ozonesonde observations were reduced from 56% to 30% in the UTLS, from 28% to 23% in the middle troposphere, and from 27% to 24% in the lower troposphere.*"

Line 444-445: My understanding is that Turnock et al. is referring to global climate chemistry models which can have more simple chemistry than the global models used here, so the same biases may not be present. Also, according to Figure 5, there is not a positive bias across the board. Can authors please adjust their wording or correct me if I am wrong on this point?

The forecast models used in chemical reanalyses incorporate a variety of complexity in their chemical mechanisms. Positive model ozone biases over major polluted regions are widely reported in various global chemical transport models (CTMs) and chemistry climate models (CCMs) (Travis et al., 2016; Young et al., 2018; Turnock et al., 2020). Like global CTMs and CCMs, some of the forecast models show positive model biases over the polluted regions as observed in TCR-2 and GEOS-Chem. Meanwhile, in other forecast models, effective optimization applied specifically for surface ozone may have contributed to the smaller surface ozone biases as observed in CAMSRA and RAQMS. Thus, we have revised this part as follows:

(p. 17, lines 508−515)
"*This finding is likely related to the limited sensitivity of satellite observations in directly constraining surface ozone. This suggests that reanalysis surface ozone biases may be largely influenced by biases in the forecast models, such as the widely recognized positive bias over major polluted regions reported in various CTM and chemistry-climate model (CCM) simulation (Travis et al., 2016; Young et al., 2018; Turnock et al., 2020). Similar to CTMs and CCMs, the forecast models used in chemical reanalyses incorporate varying degrees of complexity in their chemical mechanisms. The maturity of these forecast*

models, including effective optimization applied specifically for surface ozone, may have contributed to the smaller surface ozone biases observed in CAMSRA and RAQMS."

Lines 448-449: But the GEOS-Chem analysis showed almost zero influence on ozone, how does that square with this conclusion? As authors say, the impact of the measurements varied widely, so it's not convincing to simply state it was important without any qualifying statements.

We have revised this section and included a quantitative statement, emphasizing the significant improvement over the northern mid-latitudes, including GEOS-Chem analysis, as follows:
(p. 17, lines 518−520)
"*These OSEs suggested the importance of including precursor measurements, especially for NOx, to improve ozone analysis in the middle and lower troposphere over the northern midlatitudes, which led to reductions of 58–92% in model ozone biases relative to ozonesonde observations.*"

As for GEOS-Chem adjoint, we have added (1) more in-depth discussion on why the impact is different from TCR-2 (p. 15, lines 439−442) as mentioned above and (2) additional analysis results to show the impacts of assimilating different satellite products (DOMINO) into GEOS-Chem adjoint on ozone in the Supplementary material.

Technical corrections
Line 191: typo in "difficultiy"

This phrasing has been corrected

Data availability statement has typo.

The phrasing ", because" has removed from line 526.

Reference

Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., Yu, K., Miller, C. C., Yantosca, R. M., Sulprizio, M. P., Thompson, A. M., Wennberg, P. O., Crounse, J. D., St. Clair, J. M., Cohen, R. C., Laughner, J. L., Dibb, J. E., Hall, S. R., Ullmann, K., Wolfe, G. M., Pollack, I. B., Peischl, J., Neuman, J. A., and Zhou, X.: Why do models overestimate surface ozone in the Southeast United States?, Atmos. Chem. Phys., 16, 13561–13577, https://doi.org/10.5194/acp-16-13561-2016, 2016.

Turnock, S. T., Allen, R. J., Andrews, M., Bauer, S. E., Deushi, M., Emmons, L., Good, P., Horowitz, L., John, J. G., Michou, M., Nabat, P., Naik, V., Neubauer, D., O'Connor, F. M., Olivié, D., Oshima, N., Schulz, M., Sellar, A., Shim, S., Takemura, T., Tilmes, S., Tsigaridis, K., Wu, T., and Zhang, J.: Historical and future changes in air pollutants from CMIP6 models, Atmos. Chem. Phys., 20, 14547–14579, https://doi.org/10.5194/acp-20-14547-2020, 2020.

Young, P. J., Naik, V., Fiore, A. M., Gaudel, A., Guo, J., Lin, M. Y., Neu, J. L., Parrish, D. D., Rieder, H. E., Schnell, J. L., Tilmes, S., Wild, O., Zhang, L., Ziemke, J., Brandt, J., Delcloo, A., Doherty, R. M., Geels, C., Hegglin, M. I., Hu, L., Im, U., Kumar, R., Luhar, A., Murray, L., Plummer, D., Rodriguez, J., Saiz-Lopez, A., Schultz, M. G., Woodhouse, M. T., Zeng G.: Tropospheric Ozone Assessment Report: Assessment of global-scale model performance for global and regional ozone distributions, variability, and trends, Elem. Sci. Anth., 6: 10, doi:10.1525/elementa.265

Reply to Comments by Owen R. Cooper (TOAR Scientific Coordinator of the Community Special Issue) on:

Assessing the relative impacts of satellite ozone and its precursor observations to improve global tropospheric ozone analysis using multiple chemical reanalysis systems

Sekiya, T., Emili, E., Miyazaki, K., Inness, A., Qu, Z., Pierce, R. B., Jones, D., Worden, H., Cheng, W. Y. Y., Huijnen, V., and Koren, G.

EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-2426
Discussion started: 20 August 2024; Discussion closes 01 Oct., 2024

We would like to thank Dr. Owen R. Cooper for his comments on our manuscript, which greatly improved the quality of the manuscript. We have revised the manuscript accordingly, and the main changes are as follows:

7) An analysis was added to highlight the seasonal cycles in the improvements in ozone analysis by data assimilation (Figure 5 in the revised manuscript).
8) A more in-depth discussion was added on the impacts of discontinuity in the current satellite observing system on tropospheric ozone analysis, as well as potential strategies to address this issue after the termination.
9) An analysis was added to the Supplementary material to explain the reasons behind the large differences between TCR-2 and GEOS-Chem adjoint based on only the OMI NO_2 assimilation.

Individual comments (in black) and point-to-point replies to them (in blue) are listed below. Revised text (*italicized font*) from the updated manuscript is in quotes below.

Note that we have added line types in Figure 4 in the revised manuscript to assist readers with color vision deficiencies according to the comments from Editorial Support.

This review is by Owen Cooper, TOAR Scientific Coordinator of the TOAR-II Community Special Issue. I, or a member of the TOAR-II Steering Committee, will post comments on all papers submitted to the TOAR-II Community Special Issue, which is an inter-journal special issue accommodating submissions to six Copernicus journals: ACP (lead journal), AMT, GMD, ESSD, ASCMO and BG. The primary purpose of these reviews is to identify any discrepancies across the TOAR-II submissions, and to allow the author teams time to address the discrepancies. Additional comments may be included with the reviews. While O. Cooper and members of the TOAR-II Steering Committee may post open comments on papers submitted to the TOAR-II Community Special Issue, they are not involved with the decision to accept or reject a paper for publication, which is entirely handled by the journal's editorial team.

General Comments:
TOAR-II has produced two guidance documents to help authors develop their manuscripts so that results can be consistently compared across the wide range of studies that will be written for the TOAR-II Community Special Issue. Both guidance documents can be found on the TOAR-II webpage:
https://igacproject.org/activities/TOAR/TOAR-II
The TOAR-II Community Special Issue Guidelines: In the spirit of collaboration and to allow TOAR-II findings to be directly comparable across publications, the TOAR-II Steering Committee has issued this set of guidelines regarding style, units, plotting scales, regional and tropospheric column comparisons, tropopause definitions and best statistical practices.

Guidance note on best statistical practices for TOAR analyses: The aim of this guidance note is to provide recommendations on best statistical practices and to ensure consistent communication of statistical analysis and associated uncertainty across TOAR publications. The scope includes approaches for reporting trends, a discussion of strengths and weaknesses of commonly used techniques, and calibrated language for the communication of uncertainty. Table 3 of the TOAR-II statistical guidelines provides calibrated language for describing trends and uncertainty, similar to the approach of IPCC, which allows trends to be discussed without having to use the problematic expression, "statistically significant"

We sincerely appreciate your comments.
The color scheme specified by the guidelines is applied to the spatial maps in the manuscript. For the ozone mixing ratio unit, "nmol mol^{-1}" is used according to the guidelines. This study focuses on the year 2010 and does not analyze trends. Moreover, the expression "statistically significant" is not used.

Specific Comments:
Line 220

It's not clear which TOAR surface ozone product was used. Are you using the pre-complied 2x2 degree gridded product, available from the PANGAEA repository? If so the PANGAEA citation needs to be provided, as follows:
Schultz, Martin G; et al. (2017): Tropospheric Ozone Assessment Report, links to Global surface ozone datasets [dataset publication series]. PANGAEA, https://doi.org/10.1594/PANGAEA.876108, Supplement to: Schultz, MG et al. (2017): Tropospheric Ozone Assessment Report: Database and Metrics Data of Global Surface Ozone Observations. Elementa - Science of the Anthropocene, 5:58, 26 pp, https://doi.org/10.1525/elementa.244

Yes, we used the precompiled gridded data from the PANGEA repository and have included the citation for it in the revised manuscript.

The model/reanalysis evaluation is conducted across the full year of 2010. As reported by Logan (1999), ozonesondes show strong seasonal cycles of ozone in all latitude bands, with mid-tropospheric ozone at northern mid-latitudes increasing by more than 50% from winter to summer. Was there any seasonal dependence regarding the improvement achieved by assimilating satellite data?

Yes, we found a distinct seasonal dependency in the improvements by data assimilation as well as in the reanalysis performance, which varied among the reanalysis data sets and locations. To clearly illustrate the seasonality of the reanalysis bias and the impact of data assimilation, we have added monthly timeseries of biases in the reanalysis and control runs to Figure 5 and the corresponding description in section 3.2.2 as follows:

(p. 10, line 305–p. 11, line 308)
"*As shown in Figures 5 and S1, over the northern mid and high latitudes, RAQMS showed larger seasonal amplitudes in model bias, with a maximum in boreal spring, compared with CAMS, TCR-2, and IASI-r. Over the southern mid and high latitudes, CAMS and RAQMS showed larger negative biases in austral summer and fall compared to other reanalyses, while TCR-2 and IASI-r exhibited maximum positive biases in austral spring.*"

(p. 11, lines 312–314)
"*The seasonal variations of model biases were approximately 20% larger in IASI-r compared to the other models. IASI-r showed maximum negative biases over the northern*

midlatitudes and maximum positive biases over the southern high latitudes during summer."

(p. 11, lines 318–320)
"*IASI-r exhibited larger biases compared to other models over the mid and high latitudes in summer in both hemispheres, whereas RAQMS and TCR-2 showed larger negative biases over the mid and high latitudes in winter in both hemispheres.*"

(p. 11, line 325)
"*Seasonally dependent biases were also improved by data assimilation over all the latitude bands.*"

(p. 11, lines 331–333)
"*The seasonal variations in model biases were also reduced. However, multi-system mean biases in winter remained over the high latitudes in both hemispheres, likely due to the limited number of assimilated observations in these regions during winter.*"

(p. 11, line 339)
"*In these regions, the seasonal dependency of model biases was not improved by data assimilation.*"

The model output and the reanalysis were evaluated against ozonesondes using all observations collected in broad latitude bands, rather than comparing models and observations at individual monitoring sites. Can you comment on why this approach was chosen? I assume it's because you need a large sample size, due to the fact that most individual monitoring sites only sample the atmosphere once per week, which fails to provide accurate monthly means, as reported by Logan (1999). A new paper published in the TOAR-II Community Special Issue discusses the challenges of detecting long-term ozone trends based on once-per-week sampling and their Figure 1 shows the errors associated with trying to characterize monthly mean ozone with just 4 profiles per month (Chang et al., 2024).

Our focus was on validating ozone profiles on broader spatial scales to provide regional-scale reanalysis performance that reflects regional characteristics. This approach was chosen instead of evaluating reanalysis performance at individual observation sites, which can be affected by sparse temporal sampling, limited spatial coverages, and the influence of local-scale processes. To achieve this, validation was performed using ozonesonde observations aggregated across broad latitude bands. This methodology is supported by Tilmes et al. (2012), who emphasized that regional aggregates of individual sites provide a more representative characteristics of larger regions. Nevertheless, we acknowledge that the number of observations within each latitudinal band may not always be sufficient to fully capture regionally representative model performance, as discussed in Miyazaki and Bowman (2017). This limitation is more carefully discussed in the revised manuscript as follows. Note that temporal trends, including challenges related to limited sampling frequency, will be analyzed in companion papers from the TOAR-II Chemical Reanalysis Working Group (Jones et al., submitted).

The following sentence has been added in Section 2.3.2.
(p. 8, lines 237–244)
"*The validation was conducted against ozonesonde observations collected for five latitude bands to evaluate the global reanalysis performance in a manner that reflects regional characteristics. This approach was chosen instead of evaluating reanalysis performance at individual observation sites, which can be influenced by sparse temporal sampling, limited spatial coverage, and the influence of local processes. Aggregating individual ozonesonde sites with similar characteristics provide a more representative view of larger regions, as demonstrated by Tilmes et al. (2012). However, we acknowledge that the number of observations within each latitudinal band may not always be sufficient to fully capture regionally representative model performance (Miyazaki and Bowman, 2017) or to accurately evaluate long-term trends (Chang et al., 2024).*"

Minor Comments:
Figures 1, 2 and 3 are very hard to read because the panels are so small. The panels could be made larger if they were arranged vertically, in 3 columns; also, the color bars can only be shown once, instead of repeating for each panel.

According to your suggestion, Figs. 1, 2, and 3 have been modified as follows:
(1) Panels were enlarged and arranged vertically (3 columns x 7~6 rows).
(2) Color bars in each panel were removed and shown once below the panels.
(3) X-axis and y-axis labels were removed and shown once for each column and row

(only Fig. 1).
(4) Titles on top of each panel were removed and shown once for each column and row.

Similar changes were made for other figures according to the comments from anonymous reviewer #2.

Similarly, Figure 5 is hard to read because the panels are too small. The lat/lon numbers can be removed and the color bars can only be shown once.

We have enlarged individual panels, removed lon/lat labels and color bars, and showed the color bar once below the panels.

line 25
aircrafts should be aircraft, which can be either plural or singular

This phrasing has been corrected.

line58
please explain the perfect model assumption

We have added an explanation of this assumption as follows:
(p. 3, lines 60–62)
"*the perfect model assumption, which assumes that the forecast step within the data assimilation does not add systematic errors through model processes (Lahoz et al., 2010). Hence reanalyses inherit underlying model biases to an extent that depends on the frequency and sparseness of observations.*"

Line 217
(30°N–30°N) should be (30°S–30°N)

This error has been corrected

References:

Miyazaki, K. and Bowman, K.: Evaluation of ACCMIP ozone simulations and ozonesonde sampling biases using a satellite-based multi-constituent chemical reanalysis, Atmos. Chem. Phys., 17, 8285–8312, https://doi.org/10.5194/acp-17-8285-2017, 2017.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Conley, A., Schultz, M. G., Saunois, M., Thouret, V., Thompson, A. M., Oltmans, S. J., Johnson, B., and Tarasick, D.: Technical Note: Ozonesonde climatology between 1995 and 2011: description, evaluation and applications, Atmos. Chem. Phys., 12, 7475–7497, https://doi.org/10.5194/acp-12-7475-2012, 2012.