We thank the reviewer for their careful reading of our manuscript and constructive feedback. Our responses are given in red below.

Reviewer #1

Overall Assessment

The paper uses a dynamical linear model (DLM) to fit slowly varying smooth long-term trend lines to ozone time series from a number of sources: merged observed time series from satellites and simulated time series from two sets of multi-model simulation exercises. The obtained long-term trend lines are then used to derive ozone trends for several atmospheric layers and for different multi-decadal periods. The used data and the methodology are solid and consistent with many other publications. The derived trends are important and help to assess the expected recovery process of the ozone layer on the basis of both satellite observations and model simulations. Overall, I think this is a good and important paper and well suited for ACP.

Major Question

However, I do have a major problem with the extrapolations of trends from the 2000 to 2021 period to the end of the current century, which are presented in sections 7 and 8. To me there are so many uncertainties involved and the end results are, at best, very vague, uncertain and even doubtful. I think this entire part of the paper is unnecessary, weakens the main messages of the previous parts, and makes the paper longer and less clear. I strongly suggest to omit large parts of sections 7 and 8, especially Figs. 10 and 11 and their discussion. Ozone changes after 2030 or so depend a lot on future GHG emissions. CCMI-1 REF-C2 and CCMI-2022 REF-D2 use quite different GHG scenarios, have changes in the used models. Both model exercises make their own predictions, so there is not much point in repeating them here. Because the scenarios and models differ, and the real future emissions are likely to be different again, I don't see much point in using 2000 to 2021 changes to forecast anything in the far future. Matters get even worse when observed 2000 to 2021 changes that do not match the corresponding simulated changes are extrapolated to the end of the century using the poor correlations found for the model simulations. To me this seems a clear case of garbage in = garbage out, and an unnecessary, confusing and potentially misleading effort.

We appreciate the reviewer's concern over our exploration of a possible emergent/observational constraint on future ozone, and after reviewing the manuscript agree that the presentation could have been more careful in wording and clear in interpretation. The reviewer correctly points out that the "scenario uncertainty" associated with the uncertain future emissions (e.g. of GHGs) is a fundamental barrier to forecasting end-of-century ozone trends. However, the analysis presented in this paper is not intended to be interpreted as a forecast or extrapolation, but rather as an exploration of underlying model sensitivities based on a widely used emergent/observational constraint framework that has been applied in many contexts throughout climate research.

While emergent constraints can provide no meaningful information on which future emissions scenarios are most likely, they can identify cross-model sensitivities relating past to future behavior, and when combined with appropriate observational constraints can help narrow the "model sensitivity uncertainty" for future projections (i.e., the uncertainty associated with differing model responses to a given forcing or combination of forcings such as GHGs and ODSs, related

to the use of different models). Such observational constraints cannot reduce “scenario uncertainty”, and it was certainly not our intention to imply otherwise.

We believe that the analysis presented here constitutes a useful contribution towards understanding cross-model sensitivities to past and future ozone changes, and also highlights the limitations in using observations to constrain model behavior. Therefore, we have decided to retain this section, but have made several key revisions and modifications to clarify the interpretation and avoid any suggestion that this analysis should be read as a forecast or “extrapolation” of future ozone values. These changes consist of the following:

- Changed title to avoid the suggestion that we are forecasting future ozone. The title has been changed from:
  - “**Uncertainties in recent tropical stratospheric and tropospheric ozone changes restrict our understanding of future total column ozone change” to:**
  - **Uncertainties in recent tropical stratospheric and tropospheric ozone changes prevent the use of observational constraints for assessing future ozone change**
- We have added explicit recognition and caveating of scenario uncertainty at numerous points throughout the manuscript, including in the abstract, to emphasize that future emissions pathways represent a key unknown with regards to end-of-century ozone.

- We have added additional motivating discussion in Section 7 that explains the physical reasoning for potential emergent constraints, and why they are worth exploring.

- We have removed the word “predicted” (as in, “observationally-predicted”) from the discussion in an attempt to avoid implying that we are offering a definitive prediction for the end-of-century ozone values.

I do like Figure 8 and its discussion. Checking whether applying the DLM to the shorter (observational) period from 2000 to 2021 matches applying the DLM to the longer (simulated) period from 1960 to 2100 makes perfect sense.

We are glad to hear the reviewer found this analysis useful.

I also like Figure 9 comparing the ozone trends during the period of decline with those during the recovery period. I would like to see the same Figure for the trends from the satellite observations. Or those observed trends should also be plotted in Figure 9 (might become too crowded).

We thank the reviewer for these comments and agree that including the observational data on these plots is a useful addition. We have added the observations to Figure 9.

Minor Comments

Line 65: are the DU numbers for trends - then they should be DU per decade, or are the for twenty year changes (not trends)? As is now, I found this misleading. Please clarify.

The reviewer is correct, the values given are best referred to as “changes” (in units of DU, as given), not “trends” (e.g., in DU/decade). We agree this was imprecise and have changed the

word “trends” to “changes” when referring to changes in the DLM background (in units of DU), which is the predominant quantity we consider in this paper.

Line 72 to 74: This sentence does not make much sense to me, and captures the major problem I have indicated above. I strongly suggest to drop this sentence, and the corresponding parts of sections 7 and 8.

We have modified this sentence, see our response to the major comment above.

Line 114: I think you need to include a discussion of more recent tropospheric ozone trend updates from the TOAR II initiative. It would also be good to compare your tropospheric ozone trends to trends from TOAR II.

This is a good suggestion. The TOAR II results were not ready for our initial submission, but we have added information here on the overall results from that activity. Also, we have now included the TOAR II multi-observational mean timeseries in our analysis. Both the TOAR II results and new data set are described in a paper submitted this spring by Hubert et al.:

Hubert, D., Miyazaki, K., Dufour, G., Pennington, E. A., Sofieva, V., Arosio, C., Barret, B., Boynard, A., Coldewey-Egbers, M., Cuesta, J., Heue, K.-P., Keppens, A., Kramarova, N. A., Lambert, J.-C., Loyola, D., Orfanoz-Cheuquelaf, A. P., Siddans, R., Van Malderen, R., Veefkind, P., Wespes, C., Ziemke, J. R., Tropospheric Ozone Assessment Report II: Past and present tropospheric ozone using satellite observations, Philos. Trans. R. Soc. A. (submitted).

Line 181: one "a" too much in overlap

Fixed

Line 193: I am not sure that Petropavlovskikh et al. 2025 is really a good reference for the SBUV-COH dataset. It would be better to also refer to the SBUV-COH Webpage / URL: https://www.star.nesdis.noaa.gov/data/smcd1/ozone/SBUV_OMPS_COH/

We were originally advised by our coauthor Jeannette Wild (creator of SBUV-COH) to use the Petropavlovskikh et al. 2025 reference since no peer reviewed dataset publication exists, and other publications have used that reference. However, there is a technical note on the COH version relevant to this paper, and we have included that as an additional reference (Wild et al., 2025) and also changed the version number to reference the release date of the COH dataset version we are using (i.e., v20250612) rather than the SBUV data set version we quoted previously.

Line 238: Is that the MERRA2 zonal mean tropopause, since the satellite data sets all seem to be zonal means as well? Or are you using longitudinally resolved data. Please clarify.

Done. It is zonal mean.

Line 267: how similar are the WMO 2011 and WMO 2018 ODS scenarios (or their corresponding equivalent effective stratospheric chlorine). Can you make a statement on that?

This is a good idea. We've added the following text - "In comparison to the WMO 2010 scenario, the WMO 2018 scenario has a virtually identical peak value of equivalent effective stratospheric chlorine (EESC), but a slower decline that results in a 5 year delay in the return date to 1980 values (Lickley et al., 2018)."

Table 1: I think it would be helpful to have an additional comparable table that summarizes the different observational datasets. In all the following Figures I kept wondering which observational dataset is which.

We agree, and have added a new Table 1 showing the datasets used in the paper. In putting this table together, we realized that since submission of the original manuscript we have addressed a data issue in SWOOSH, necessitating an update from v2.72 (i.e., the version used in the original submission) to v2.73 (used in the revision). This update includes removal of UARS MLS O3 from mid-1997 to the end of its record (1999) due to the failure of the radiometer used for pressure/temperature retrievals, which impact ozone (Livesey et al., 2003). The UARS MLS O3 data have a high bias/drift in the mid stratosphere during this period, which causes the DLM fit to be too high in 2000 and causes post-2000 SWOOSH trends to be less positive than they should be. The manuscript has been updated to include the latest SWOOSH data, and the text has been updated in several places to reflect these changes.

Figure 2: It would be good to have a third panel that plots the CESM1-WACCM time series also just for the 1984 to 2021 period. Or plot the model results in the top panel along with the SWOOSH time series.

The goal of this figure is not to compare WACCM and SWOOSH, but rather to give the reader a qualitative sense for what the DLM fits look like and what the important considerations are. Rather than add yet another panel or clutter the figure by combining them, we've modified the text to try and make the figure's purpose clearer (see also response to comment by reviewer #2).

Figure 9: As mentioned above, it would be good to also compare the two trend periods for the satellite observations. Either also plotted in Figure 9 (might become too crowded), or in separate Figure.

We agree, see response to similar comment above.