

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., Orfanos-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.

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

Reviewer #2

Overview

This article presents recent stratospheric and tropospheric ozone trends in the tropics and investigates the agreement between satellites and CCMs. The authors use various satellite products and compare 2 model experiments. They estimate model and satellite data trends with a dynamical linear model and found large spread in the trend results. They conclude that satellite observations do not agree well with each other, which makes the model evaluation difficult.

Furthermore, the authors present long-term ozone changes until 2100 and evaluate model differences for that period. They also investigate the effect of different observational periods on the trend results and evaluate the ratio between depletion and recovery trends.

Finally, they analyze correlations between early and late 21st century trends to assess how the disagreement in observations affects the understanding of future recovery. They found that the large spread in observations propagate in the model predictions, and that there is no clear observational constraint that could be used to predict future ozone changes.

The manuscript presents a comprehensive analysis of tropical ozone trends and investigates current discrepancies between satellite observations and model estimates, which is an important contribution to the field of research and fits to ACP. Overall, the manuscript is well structured, with room for some improvement as suggested in the following comments:

Major comments

- Figures 3+4: Even though these figures give a nice overview over the data spread, they are very crowded and it is almost impossible to distinguish the lines and thus very hard to interpret them. There is just too much information in those figures. Please ask yourself what the main aim of the figures is and think about how to present the main message to the reader. If the main aim of the figures is to show the spread amongst the datasets, I suggest the following revision (2 options):
 - Option 1 (my preference):
 - Keep the colored model lines as they are (with thinner linewidth though) but remove the observational data in Fig. 3 and Fig. 4 and add an additional figure with the observational datasets (with differently colored lines).
 - Use then the same new observational color code also for the Kernels in Fig. 5 and 11
 - Show the spread of the observational data with a semi-transparent shading in Figs. 3+4, and the spread of the model data with a shading in the new observational figure, to make comparison between model and observations possible.
 - Option 2 (if you really want everything in a single figure):
 - In each panel, show all models with thin, semi-transparent lines in one color, and all observational lines with thin, semi-transparent lines in another color.

- If desired, you could highlight single lines in each panel (e.g. some of those that you mention in the text), for example by putting them in bold and by labelling them directly at the right edge of the lines
- If desired, you could add a bold line representing the mean of the models and the observations respectively
- I would suggest to use additional colors also for the observational datasets in the right panels. You could also label the bars with vertical labels on the x-axis, but it might be difficult to integrate that into the overall layout of the figure.
- Generally, I think that the lines can be thinner to increase visibility
- I see that you plot for each model the available realizations, which is not stated in the caption (the reader might wonder why there are several lines with the same color).

We thank the reviewer for this comment, and accept that Figures 3-4 as originally devised were very crowded, which made it difficult to distinguish the various observational data sets and models. As the reviewer correctly points out, one of the main aims of these figures is to show the spread among the models and data sets.

After considering the reviewers ideas and trying out several ways of modifying these figures, we have landed on a new layout that separates the timeseries plots and the 2000-2021 trends into two figures, which replace the original Figures 3 and 4. The new layout addresses the reviewers readability concern while both preserving the essential information in the original figures and providing several advantages over the original figure layout. For example, in the first figure (new Figure 3), we display timeseries from both model experiments side by side. This allows the reader to quickly see differences between the two, whereas previously one would have to look across two different figures on different pages. Similarly, in the new trend figure, the reader can compare both model experiments to one another. In the new trend figure, we have also grouped the models by family so that changes across the generations/experiments can be seen easily.

- Section 5 relies heavily on the degree of agreement between trend estimates (overlap coefficients), but the coefficients are not given in any figure or table. Please provide those values, for example in a correlation matrix (ideally colored with a heatmap function). I think such a representation would simplify the interpretation of the results and make it easier for the reader to follow.

This is a good idea to provide this information, although since we are dealing with 2 multimodel experiments and ~7 observational data sets in each of 6 atmospheric layers, providing the suggested heatmaps involves 6 ~9x9 panels, and we didn't feel that quantity of information is an appropriate use of space within the main manuscript. Therefore, we've generated the heatmaps and have placed them in a supplement (Figs. S1-S6), for the interested reader who might find them useful. Also, we would like to note that our original Figure 6 neatly conveys the information that is discussed quantitatively (in terms of overlap values) in Section 5. In addition, we have also included a table of mean value changes in the supplementary information (Table S1).

- Section 7.3 and section 8 rely on the assumption, that late-century trends are correlated with the corresponding early-century trends (observational constraint). In the first paragraph of Section 7, you present the key questions that you want to answer, but I am missing some discussion about the physical reasons for a constraint-assumption. For example, why should high early-century ozone trends always lead to high end-of century

ozone trends? In principle, there could be a change in trend-intensity over time. Please comment on that.

We now recognize that we did not properly motivate this analysis, and agree that it would be helpful to have a discussion upfront about the potential physical/chemical processes that could lead to correlations between early and late century ozone trends across different models. We have therefore added text at the beginning of section 7 that provides the motivation for this analysis and gives a few examples of the potential mechanisms that might give rise to the correlations.

General comments

- Section 4.2 DLM example: in principle it is a nice idea to show an illustrative example of the DLM fit to the data, but the section is very short and the the purpose is not clearly stated.
 - What do the authors want to show with the figure?
 - Why is only the solar cycle shown and no other predictors?
 - Please expand the text of Section 4.2 accordingly

We agree that this section was too short and not fully explained, and have expanded the discussion. The goal of this section/figure is simply to show a couple of example DLM fits to give the unfamiliar reader a qualitative sense for what they look like, and to highlight the importance of the solar cycle predictor.

- In some parts of the result sections, the authors give methodological details. The result sections would be smoother if those parts would be moved to the earlier methods/data sections:
 - move DLM-explanations such as Kernel description (l. 378-383) to DLM section (Section 4)

Done

- move the explanation of overlap coefficients (l. 401-404) to the previous paragraph, because it applies to all the following subsections of Section 5

Done

- You compare 2 model experiments, CCMI-1 (with REF-C2 simulations) and CCMI-2022 (with REF-D2 simulations). Please be consistent when referring to those models/setups in the text. Sometimes you refer to CCMI-1 vs. CCMI-2022 experiments (e.g. l. 396), sometimes you use the full name CCMI-1 REF-C2 vs. CCMI-2022 REF-D2 (e.g. l. 414 or l. 360), and sometimes you refer only to REF-C2 vs. REF-D2 (e.g. l. 440). This leads to confusion. Please define in section 3 (CCM description) clear names that you will use throughout the whole manuscript.

We agree that the use of various combinations of the project (e.g., CCMI-1) and experiment (e.g., REF-C2) was confusing, so we have revised the manuscript to mostly refer to the specific experiments like REF-C2 and REF-D2. We note that in some cases, e.g. when referring to more than one experiment within a project (e.g., when referring to both the fGHG and REF-C2 experiments in CCMI-1) or characteristics of the project as a whole, we have retained reference to the project name.

Specific and technical comments

Introduction

- Line 76: The first sentence should introduce the broad context of the publication. Even though your first sentence gives important scientific information, it is very long, combines too many ideas and highlights secondary factors (e.g. details about clouds). Please revise.

We agree that in our effort to be technically accurate we lost the forest for the trees here, and have revised this sentence to try and make a simpler point that TCO matters. We have also reworded a few parts of the introduction to try and better motivate this work.

- Line 85 "During this period since ~2000": please reformulate,
Done
- Line 98 This sentence is not very clear, please revise
Done
- Line 112 "growth of the region": Not clear what is growing, do you mean population increase? Please reformulate.
Done
- Line 120-122: the percentage values are not consistent with the values given in Fig. 1, please correct
Thank you for catching this. Done.
- Line 137 "span too wide a range": please reformulate, e.g. "we demonstrate in Sect. 8 that the current range of observational trend estimates is too large to effectively serve...."

Done

Sect. 2

- Line 167, line 173, line 182: repetition of "we also use/include", please vary.

Done

- Line 185 "2004 and the 2024": remove "the"

Done

- Line 205 "Note, ...": "Note that ..."

Done

Table 1

- Improve the caption (e.g. CCMs used in this study to simulate ozone from 1960 to 2100)

Done

- CNRM-MOCAGE: Does put the "" to the name here, not to the number of realizations

Done

Sect. 4

- Line 325: confusing sentence structure, move "and detrending": change to "... we compute the ENSO index by averaging and detrending the surface temperatures from each simulation..."

Done

- Line 346 "shorter time periods such as the observational record": Reformulate. Shorter than what? Which observational record?

We have re-worded this section to be more clear

- Line 348 (caption Fig. 2) "tropopause – 1hPa": change to "tropopause to 1hPa"

Done

- Line 349 (caption Fig. 2) "black lines with 1 sigma uncertainty shading": there is no shading for the lower black line, even though it is mentioned in the caption and given in the legend

We've updated the figure caption to reflect that the uncertainty shading is only for the observational data in this figure. As discussed in the text, there is no (statistical) uncertainty in the model PCO.

Sect. 5

- Line 353: please improve the title of section 5, "observational period" is very vague, please adapt, e.g. to "Tropical ozone changes from 1980 to 2021"

The differing satellite data sets have different start dates so we prefer not to include a date range in the title of this section. However, we have changed the text to state "satellite observational period" to clarify the type of observations to which we are referring.

- Line 378: move KDE description to Section 4 (see general comment)

Done

- Line 401-404: move to previous section (see general comment) and provide a correlation matrix there (see major comment)

Done

Sect. 6

- Line 519-522: please revise, it is not clear what the authors would like to say with this paragraph.
 - Line 519: What does "relatively negative" mean? Do you mean smaller trends?
 - Line 520: the upper stratospheric trend of ACCESS_CM2-Chem is not negative

In this sentence, we meant "relatively negative" compared to the distribution of REF-D2 model trends in the middle and upper stratosphere. The point we are trying to make is that the ACCESS whole-stratospheric trend being a (negative) outlier in Fig. 6h is due to its trends in the middle and upper stratosphere, not its trend in the lower stratosphere (which is pretty much middle of the pack). We are aware that the ACCESS upper stratosphere trend is positive, but it is on the low end of the model spread, hence the use of "relatively negative", which applies independently of the sign of the trend being discussed. All of this said, we agree that the sentence could be clearer, and have revised it to remove the use of "relatively negative".

- Line 521: remove "very", which is an unnecessary and subjective intensifier
Done
- L 523 "somewhat different": reformulate
Done
- Line 525 "dramatic increase": remove "dramatic", which is a non-scientific intensifier

Done

Sect. 7

- Line 545: If I understand correctly, the full period is used as an input to DLM, but the trend is then determined for the same period as for the first estimate (1985-2021) to compare both. Please clarify that. It reads as if the 1985-2021-trend is compared with the 1960-2100-trend, which does not make sense in this context.

The reviewer is correct in their understanding. We have modified this sentence to remove any confusion on this

- Line 549: remove "very", which is an unnecessary intensifier

Done

- Line 561 (caption Fig 8) "significant correlations are in bold": In this figure, all correlation coefficients are significant and thus bold, which makes this claim confusing here. It makes sense only in the following figures (Fig. 9 and Fig 10).

Done. This was removed and replaced with a note that all correlations are significant.

- Line 578 "a faster inorganic chlorine (Cly) increase before, and a faster decrease after, the peak Cly in the late 1990's": please remove the commas, otherwise the sentence is confusing

Done

- Line 581 "(dotted line, Fig. 9a),.": remove comma

Done

- Line 593 "multimodel-mean value of the Cly recovery/depletion trend ratio calculated over the same latitude (30° S–30° N) and vertical region from the REF-C2 experiment.": Figure captions should not contain new information. Please explain how the dotted lines are calculated also in Sect. 7.2.

Done

- Line 610: BDC is not defined

Done. It is now defined in the introduction section (fourth paragraph)

Sect. 8

- Line 676: the observations and model distributions show similar widths in the middle stratosphere, but in the upper stratosphere, the observational widths are larger (even though less large than in the lower stratosphere). Please mention that.

This is a good point. This section has been reworded.

Conclusion

- Line 730: please provide references for the negative lower stratospheric ozone trends.

Done

- Line 745 and 747 "These results highlight": repetition, use e.g. "Ultimately, the results indicate ..."

Done