Referee #1

Dear Referee,

We are grateful for reviewing our manuscript and your valuable comments and suggestions, which will largely improve our manuscript. Below you can find responses to each of your comments (marked in blue):

General comments:

The paper addresses the very interesting topic of stratospheric ozone trends and variability after 1985, split in two periods (i.e., ozone decline and recovery), with a focus on the lower stratosphere where a continuing decline of ozone in the lower tropical and mid-latitude atmosphere has been the subject of recent works. The data used in the paper are derived from Chemistry-Climate Model simulations, and a comparison of trends derived from observations composites and reanalysis data from different sources is performed.

The subject is appropriate to the Journal and contains significant original material.

The paper is generally well organized, but there are parts that need to be revisited for sake of clarity of meaning.

I recommend publication after minor revisions.

Specific comments below:

1. **Section 3 (description of the DLM approach)**

This section, especially in the last paragraph (line 182 and below should be revisited to clarify the meaning. It would be better if you split this paragraph in two parts (from sentence beginning with “It is important to note…”) so that you may explain better how the trends were calculated (“..calculated from the DLM output. ..”).

Thank you for your comment. This part has been largely rewritten for clarity and understandability as follows:

“The remaining proxies show the average correlation coefficients between -0.2 and 0.2, which is considered to be a “weak” mutual dependency.

It is important to note that in this study, all DLM calculations using model data as well as ozone composite and reanalyzes were performed for the entire 1985-2018 period. The evolution of the predicted variable is characterized in DLM by the so-called “trend-term” or background level. In simple terms, it is the evolution of a variable filtered from the known part of its variability, induced by external forcings, which are represented in DLM by proxies. The state of the predicted variable for the one-step-ahead is predicted by the Kalman filter, and the Kalman smoother is used for the marginal probability distribution of the state. The Markov Chain Monte Carlo (MCMC) method (Alsing, 2019) is used to infer the posterior distributions of the background level. Then, 200 samples of the background level were

drawn from the DLM states, which describe the posterior distribution uncertainty. It was done to determine the standard deviation of background level between these samples used to calculate the statistical significance of the results. Afterwards, the ozone trends per decade are calculated separately for the phases of the ozone evolution ([between 1985 and 1997] and [between 1998 and 2018]) from the mean background level by applying the conventional linear regression ($\alpha x + \beta$). In linear regression, α means a slope term, and, hence, a trend per decade at each grid point (latitudes x heights) are calculated as: α*length of month-to-month time series/number of decades. Finally, the statistical significance is determined for each ensemble member using the mean background level and its standard deviation between MCMC samples by the Student's t-test. The ensemble mean trends of the SOCOLv4 experiment are calculated as an average of the trends from all individual ensemble members. The statistical significance for the ensemble mean trends is calculated using a standard deviation of trends between individual ensemble members. Trends in BASIC ozone composite and reanalysis data sets were calculated using the same methodology but applying observed proxy variables in DLM, the same as presented in Ball et al. (2018) as it was mentioned above. “

Was the same approach followed for the reanalysis datasets? How was it done?

Yes, the methodology is the same. However, to analyze the ozone from BASIC observational composite and all reanalyzes, DLM uses observed proxy variables, as in Ball et al. (2018).

Please write a line also to comment on the assessment of significance.

The statistical significance of the trends was calculated using the Student’s t-test. It was done for all individual ensemble members of the SOCOLv4 reference experiment as well as BASIC observational composite and reanalyses using the sample mean background level and its standard deviation between MCMC samples. In the case of the ensemble mean of the SOCOLv4 reference experiment, statistical significance was calculated using the ensemble mean trend and its standard deviation between ensemble members of the experiment.

In text we included this as follows:

“Finally, the statistical significance is determined for each ensemble member using the mean background level and its standard deviation between MCMC samples by the Student's test. The ensemble mean trends of the SOCOLv4 experiment are calculated as an average of the trends from all individual ensemble members. The statistical significance for the ensemble mean trends is calculated using a standard deviation of trends between individual ensemble members.”

In the same paragraph, lines 181-182, please explain the “within 0.20-0.25”. Are the correlation coefficients always positive for all variables? Or you mean something different?

Thank you for the comment, it is, indeed, not so clear what is meant by “within 0.20-0.25” here. In fact, Figure 1 shows the correlation coefficients between proxies for model data.

They are generally between -0.2 and 0.2. In the text, this line has now been rephrased to be clearer.

2. **Section 4. 2 (Partial and total column ozone evolutions for the 1985-2018 period)**

This section should be re-written for a clearer meaning allover.

You start in line 222 with tropospheric ozone, which is then dropped (and revisited later for the observations), and in the same line (223), immediately after the end of the sentence you note changes in mesospheric ozone.

Thank you for your comment. This paragraph has been fully rewritten and now evolutions of ozone from all used data sets are described layer-by-layer so that readers can easier follow the text. The revised paragraph is below:

Yet, the upper stratospheric ozone evolution on the near-global scale in MERRA-2 and ERA-5 seems to be biased against BASIC and SOCOLv4 and doesn't reflect the evolution of hODSs, with the hODSs-induced ozone decline not appearing at all in MERRA-2 and being only scarcely presented in ERA-5.

"…It is worth noting that the mesospheric ozone evolution generally resembles the one in the stratosphere since the lower mesosphere in the model is still affected by the stratospheric ozone. In addition, mesospheric ozone does have a tiny positive contribution to the enhancement of the total column ozone; it demonstrates a positive trend of ~0.03 DU between 1998 and 2018.

The upper stratospheric ozone from both SOCOLv4 and BASIC demonstrates a pronounced decline during the ozone depletion phase, with a minimum after the Pinatubo eruption due to additional chlorine activation on volcanic aerosols. After 1998, the upper stratospheric ozone from both SOCOLv4 and BASIC began to increase distinctly. Yet, the upper stratospheric ozone evolution on the near-global scale in MERRA-2 and ERA-5 seems to be biased against BASIC and SOCOLv4 and does notn't reflect the evolution of hODSs, with the hODSs-induced ozone decline not appearing at all in MERRA-2 and being only scarcely presented in ERA-5. Meanwhile, ozone evolutions in the upper stratosphere during the recovery phase in MERRA-2 and ERA-5 are more reasonable but still hardly resemble those from BASIC and SOCOLv4.

In the middle stratosphere, SOCOLv4 shows a pronounced increase in ozone for a few years after the Pinatubo eruption due to an additional heterogeneous reactive uptake of N2O5 onto sulfuric acid particles, leading to a suppression of the NOx catalytic ozone cycle (Prather et al., 1992, Solomon, 1999, Rozanov et al., 2002). BASIC does not show a similar increase in ozone, which might be because the effect of the Pinatubo eruption was filtered from the data due to some reported problems in satellite ozone retrieval during strong stratospheric aerosol loading (Davis et al., 2016; Ball et al., 2017). Unlike SOCOLv4, MERRA-2 and ERA-5 show a continuous decrease in middle stratospheric ozone until 1998 in MERRA-2 and around 2003 in ERA-5. In ERA-5, the start of ozone recovery due to the

hODS decrease is biased relative to the BASIC and SOCOLv4 towards the beginning of the 2000s.

In the lower stratosphere, the ozone decline is seen in all presented data sets for the ozone depletion phase. The ozone minimum in this region has been visible for a few years after 1991 because of the enhancement of chlorine activation after the Pinatubo eruption (Solomon et al., 1993), after which ozone starts to recover. The LSO evolution in MERRA-2 shows a decline similar to SOCOLv4, while in ERA-5, the ozone decline is much stronger than in all other data sets. Starting from 1996, the LSO evolution is highly variable as a result of natural unforced variability. The extra-polar LSO evolutions from SOCOLv4 fluctuate around zero. BASIC shows the continuous decline of extra-polar LSO throughout the whole period, which is estimated to be ~2DU since 1998. In ERA-5 and MERRA-2, ozone evolution has completely opposite behavior during the last years of the period, showing either decline or increase, correspondingly. MERRA-2 and SOCOLv4 generally match each other and only a few years toward the end of the period MERRA-2 show a pronounced decline. ERA-5 demonstrates a much stronger increase in ozone during the ozone recovery phase, but the ozone starts to increase later than in other data sets and is virtually biased against BASIC and SOCOLv4.

Extra-polar tropospheric ozone in both SOCOLv4 and ERA-5 are similar, showing a pronounced increase throughout the whole period of ~4DU/period, which is facilitated by an increase in the number of ozone precursors, mainly NOx (see Figure 2) and CO. MERRA-2 in the troposphere is not shown because it dramatically disagrees with expectations and other data sets. ERA-5 is more applicable in the troposphere, showing a general agreement with SOCOLv4.

Overall, the total column ozone evolution modeled with SOCOLv4 is well within the range of evolutions from other data sets, for which the total column ozone is available. The total column ozone evolution simulated with SOCOLv4 agrees better with MSRv2. For ERA-5, we see a much stronger increase in total column ozone than in the other data presented resulting from the abnormal ozone evolution in the stratosphere. MERRA-2 shows the marginal decline of total column ozone compared to SOCOLv4 and other reanalyses, which may also be due to some flaws in this data, primarily for the stratosphere. Nevertheless, the presented time series display a general increase of the extra-polar total column ozone. Yet, based on the obtained results, the important message is that the ERA-5 and MERRA-2 reanalyses still do not fit well for ozone trend analysis. We cannot demonstrate the source of this discrepancy, but it could be related to some issues with their underlying models, specific observational data processing, or assimilation approaches. We can be confident in saying that this is not related to DLM, as the evolution of ozone from these reanalyses prior to the application of DLM (dashed lines in Figure 3) also exhibits this anomalous behavior. Yet, the MSRv2 reanalysis behavior is much smoother and more understandable, which agrees rather well with the SOCOLv4 simulations."

Please reorganize the paragraphs, so that you facilitate reading and clarify your findings, especially the comparison to reanalysis. It is in this section you need to justify the statement in **Section 5** (line 313 -) "...We also show that the MERRA-2 and ERA-5 reanalyses are less suitable for ozone trend analysis because..."

Thank you for this comment. Indeed, in Figure 3 it is seen that ozone evolutions from both MERRA-2 and ERA-5 reanalyzes are generally biased in the stratosphere against BASIC observational composite and SOCOLv4. We cannot demonstrate the source of this discrepancy and it is out of the scope of this paper, but it might stem from some issues with their underlying models, specific observational data processing, or assimilation approaches. This has now been properly addressed in the rewritten paragraph (see the rewritten paragraph in just above comment).

The same comment for reorganization of paragraphs applies to all remaining sections, as it might even be confusing at some points.

Agreed. The next sections have also been restructured to have the description layer-by-layer (see reformulated subsection 4.3. Simulated and observed long-term ozone trends for the decline and recovery phases and section 5 Discussion and summary.)

3. The statement in the last lines of **Section 5** (line 350 -)

"The results further confirm the poorly understood ongoing decline of ozone in parts of the extratropical lower stratosphere…"

What do the results confirm? That there is a continuous ozone decline in the extratropical lower stratosphere (statistically significant), the origins of which are poorly understood, or that there are indications of an ozone decline, with a patchy response, with origins that are not understood?

So please rephrase to clearly present your findings and their importance.

Rather the second. We rephrased this part as follows: **"**The results further confirm that there are marginally significant negative ozone changes in parts of the low latitude lower stratosphere. This result agrees in general with the negative trends extracted from satellite data composite, however the simulated magnitude and significance are lower than in observations."

Technical and other comments

Abstract

Line 9: "…derived from observations and reanalyses." Please refer here to the datasets, e.e. "namely the BASIC composite of ozone, and…"

Agreed. It has been reformulated as follows:

"We applied the new Earth system model SOCOLv4 to calculate long-term ozone trends and compare them with trends derived from BASIC ozone composite and MERRA-2, ERA-5, and MSRv2 reanalyzes."

Line 14 "…do not agree with some observation composite analysis." Which ones? Specify

This sentence has been modified: "….with BASIC ozone composite."

Introduction

Line 30 "… no or a …" change to "none or a …"

Done.

Line 69 "…BDC…" please give the name in full before using acronym for the first time, the reference to Butchart et al could be given here as well.

Agreed. It has been rewritten like this: "…the meridional Brewer-Dobson circulation (BDC, Butchart et al., 2006)….".

Section 3. The description of the DLM approach

Line 168 "…are used to represent trends…" what is the meaning of these last words? Please clarify.

We agree that it was unclearly formulated. We amended it like this: "…are used to explain the variability in total/partial ozone."

Line 184 "…output DLM output." Please delete the first word

Done.

Section 4.2

For clarity, in line 219 please write "and total column ozone (represented in Figure 3 by the entire model atmosphere, upper right panel)…" or something similar.

Done.

Line 236-237 "… might be due to …" --> "… might be due to either…or…"

Done.

Section 4.3.

Line 256 "In this study we applied … and afterwards ozone changes..." please change to "In this study, we first applied… and then computed ozone changes…"

Done.

Line 261 "However…" Why "however"??

Agree, it is unneeded here and has been deleted from the text.

The same for lines 299 -300, you use “However…” twice

Yes, the one of them is deleted.

Section 5

Line 308 “…various reanalyses.” -> “… a number of available reanalyses.”

Done.

Line 318 “…the model provides…” -> “the model shows”?

Done.

Line 351 “…and show that also an ensemble approach…” could be better as “… moreover show that even an ensemble approach…”

Done.