Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2019-734-RC2, 2019

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Inconsistencies between chemistry climate model and observed lower stratospheric trends since 1998" *by* William T. Ball et al.

Anonymous Referee #1

Received and published: 27 October 2019

The paper 'Inconsistencies between chemistry climate model and observed lower stratospheric trends since 1998' by Ball et al. discusses recent ozone, stratospheric water vapor (SVW) and temperature trends in the lower stratosphere within the tropics and mid-latitudes. One conclusion is that most CCMVal2 models (in particular the multi-model mean) cannot reproduce observed trends in ozone from 1998-2017, while being able to capture temperature trends over the same period. They argue that this is only possible due to offsetting biases in simultaneous modeled SVW trends. As another important point, the authors argue that some models are better at capturing mid-latitude ozone trends inferred from observations than others, which in turn appears to be related to how lower stratospheric isentropic mixing is modeled in each case.

Without a doubt, the authors address an interesting but also highly complex topic. Their analysis therefore also requires particular care and has to be put into the context of the vast associated uncertainties. This makes it very difficult to study ozone and other trends over such short periods of time, which, in turn, links back to some weaknesses in their methodology and datasets used, which need either to be addressed or at least clearly highlighted to raise more awareness around them. Some of these challenges are already discussed in the paper, especially towards the end. However, currently, these uncertainties are not sufficiently reflected in the abstract, for example.

Major comments:

- The most concerning aspect are potential robustness issues: the authors consider very small trends that may or may not be due to actual climate change/MPA trends or simply artefacts of internal variability. On top of that, the observations are subject to uncertainties and the trends are also calculated differently (and the data preprocessed) for observations and models. At least this is how I understand section 2.2. The method to calculate the trends is also approximative. Overall, this implies that the main results might well arise from complex error propagation that for me as a reviewer is difficult to see through. This does not mean that the results may not be interesting or worthy of being published as a point of discussion. However, I also feel that some statements in the paper would ideally be tuned down and these uncertainties reflected appropriately and discussed more extensively. In particular, given the uncertainties and different methods to estimate internal variability contributions for models and observations, I have doubts about how well the 'accelerations' (second-order derivatives) in Figure 1c/d can actually be compared and how robust such a comparison can be.
- Is the singular attribution to SVW not too simplistic? Could dynamical heating not also play a role? Dynamical heating can also be quite different from what happens in the real world. In general, I consider ozone, temperature and dynam-

ical trends a coupled problem, where cause and effect are difficult to distinguish. Would lower SVW trends not also be strongly influenced by model differences in isentropic mixing for example? How about differences in radiative transfer codes?

- The use of CCMs with SSTs different from the ones from observations makes me doubt if we can at all expect the models to perform similar to observations over this short time period. If, as a result, the DLM analysis is carried out differently, can we at all expect the same results (which will depend on these aspects of variability)? From the current text, this is at least not sufficiently justified. Do all CCMs actually use different /the same SST fields? Would we expect models (or subsets of them) to be consistent in terms of SST variability, which is surely connected to lower stratospheric ozone variability due to well-known effects of the ENSO etc?
- In the same vein, the use of a single radiative transfer model for the FDH calculations is necessarily imperfect, as different radiative transfer schemes themselves will contribute to the temperature trend differences among models.
- SVW trends can be very different for CCMs (see your own Supplementary figures). Re your trends in Figure 1b: how would the same trend for SWV look like of you took the model median, or plotted trends for all individual models? Would you still come to the same conclusions?
- How does the effective vertical range for lower stratospheric ozone trends compare for models and observations? How is lower stratospheric ozone defined in terms of the vertical range covered? Could different vertical dataset resolutions play a role in the differences you find?
- Re all trends you show: given that the models have different SST fields: would we really expect the MMM to be able to reproduce historic trends? Would we not

better ask if any of the ensemble members in the multiple CCM runs can reproduce the observed pattern of lower stratospheric ozone decreases? You might argue that models with multiple ensemble members might be consistently offset from observations (your Supplementary). However, did those different ensemble members actually use substantially different SST fields? Could a lack of skill in modeling SST variability in the first place be responsible for the apparent inability of models to capture lower stratospheric trends, i.e the biases are introduced somewhere else in the system unrelated to chemistry and stratospheric dynamics? If one ensemble member can reproduce historic trends, is it then in the realm of possibilities in the modeling world to reproduce observed trends, so to speak? If yes, can you still come to such strong conclusions concerning the models' skill to reproduce past trends?

Other comments:

- l.9: 'an increase'
- l.96-109: see above. Do we expect the MMM to be able to reproduce such a short period of time on average, or are we looking for individual ensemble members for this?
- l.138: typo?
- l.179: any particular reason why you used those 12?
- Figure 1: why are units/scalings for (c) and (d) not consistent?
- l.235-255: I am still not convinced that the MMM should agree with the single-member observations. Do you have any theoretical justification for that?
- l.283: typo

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-734, 2019.