Review of “Evaluating reanalysis representations of climatological trace gas distributions in the Asian monsoon tropopause layer “ by Wright et al.

Summary

The aim of this work is to assess the performance of five different modern meteorological reanalyses in terms of their ability to reproduce the characteristic seasonal variations in UTLS trace gas composition (H2O, O3, CO) associated with the Asian summer monsoon. The core of this work is an intercomparison of H2O, O3 and CO distributions with AURA MLS observations in the area associated with the ASM. The second objective of the work is to relate and explain the evaluated trace gas concentrations and their variability in the models to different mechanisms (dynamics, physics, assimilation).

The analysis shows that the selected meteorological reanalyses are in principle able to represent the typical seasonal changes in H2O, O3, CO. It also shows substantial differences between individual models in the trace gas balance, which is a relevant result for the scientific community using this model output.

The topic of this work is well in line with the subject areas of ACP.

The data analysis seems adequate to me, but I have some general and specific comments that could help to improve the structure of the scientific paper and increase its comprehensibility. I recommend this work for publication, subject to revision/consideration of these comments.

General comments

1.

P11 Figure 3: The water vapor concentration drops across the UTLS by several orders of magnitude with highest water vapor in UT. So, I wonder whether the signal (overestimation of PCWV by all models) that you see in the top face is purely dominated by the troposphere. It would be interesting to look at the PCWV for H2O (and for the other trace gases) also in a narrowly defined layer around the tropopause level, and separately in the UT and in the LS.

2.

Chapter 1: I would appreciate to get more context on what is known on the accuracy/performance of the reanalyses with respect to these trace species in the UTLS. Are there well-known substantial H2O/Ozone/CO biases? Are these biases expected to be stronger in the ASM?

3.

Chapter 4: This chapter should be revised using a clearer structure, simpler language, by creating a logical connection to chapter 3 in order to give the reader more guidance and make it easier to follow.

In particular, more context (especially at the beginning of the chapter) on physical/dynamic tendencies would help the reader understand why you are using them (based on the findings in sect. three) and how to interpret them.

The authors could also consider adapting the chapters heading ‘Water vapour and ozone budgets’ and linking it directly to the tendencies.

4.

P19: Figure 9: The use of CERES EBAF data (Fig. 9) appears somewhat out of nowhere here (also in sect. 2.2, P5 ll. 145). Although this figure is interesting and supportive for your conclusions (regarding the overestimation of deep convection by MERRA-2), you can consider whether it is sufficient to append it to the supplement. Here, the OLR difference (CERES-each model) would also be interesting here to demonstrate also differences in OLR between CERES and the other models.

Technical corrections

P2: ll. 30-31: What exactly is meant by ‘relatively’? In general, this term is used very frequently in the paper and should be specified -if possible.

P2: ll. 32-33: “*Convection also leads to relatively low concentrations of ozone and high concentrations of CO and other pollutants (…)”.* Please specify which altitude region you are referring to (UT, LS, both?)

P2: ll. 41-42: “*at these altitudes*”*.* Do you mean the UTLS in general here?

P4: ll.80-82: What is the vertical resolution/number of levels in the various models in the height range relevant for this study? This could be interesting information for the reader. A comment and/or a corresponding addition to Table 1 would be helpful in this regard.

P5: ll. 125-125: *“subtracting forecast specific humidities (before data assimilation) from analysis specific humidities (after data assimilation)”*. To avoid misunderstandings, it may be useful to mention that data assimilation increments are the difference between the analysis and the background forecast (i.e., the first guess).

P5: ll 135: *“upper troposphere and stratosphere (p < 316 hPa)”* Is there also an upper boundary of MLS data in the lower stratosphere?

P5: ll. 142: *“tropopause based on the World Meteorological Organization (WMO) definition”*. Please clarify that you are referring to the cold point tropopause? Could you add a reference for its WMO definition?

P6: ll. 146: Can you also comment on the vertical resolution of the different AURA MLS products? Is that different for the different trace species?

P10 ll. 210-211: Is there a reason why you chose the lapse-rate tropopause here and not the cold-point tropopause (which you use in the rest of the paper)?

P10 ll 219: *"Although the reanalysis profiles are considerably warmer than indicated by Aura MLS…"* Please rephrase, for example: Although the reanalysis temperature profiles are considerably warmer than indicated by Aura MLS at all altitudes within the UTLS…

In the following cold bias in MLS is emphasized only. It should be also mentioned, that reanalyses are also affected by biases in the UTLS region. For ECMWF IFS (ERA5 based on IFS cycle 41r2), for example, there is a well-known warm bias at the tropopause (e.g., Ingleby 2016; Ingleby et al., 2017), and a cold bias in the LS radiatively which is associated with a collocated moist bias. Ingleby (2016) indicates that the warm tropopause may to some extent be caused by insufficient vertical resolution.

P11 Figure 3 (GENERAL COMMENT): The water vapor concentration drops across the UTLS by several orders of magnitude with highest water vapor in UT. So, I wonder whether the signal (overestimation of PCWV by all models) that you see in the top face is purely tropospheric. It would be interesting to look at the PCWV for H2O (and the other trace gases in other figures) also in a narrowly defined layer around the tropopause level, and separately in the UT and in the LS.

P12 ll. 245-246: "*The small biases in the regional anomalies based on these reanalyses establish that moist biases relative to Aura MLS are hemispheric in scale and are not specific to the monsoon region*" What exactly do you mean by that?

P14 ll. 285-286: In my point of view, the agreement between CAMS and Aura MLS CO distributions is pretty strong, while there is a substantial offset in MERRA-2 CO. Hence, I would recommend to make these differences a bit more prominent.

P15 ll. 300: I would recommend to delete "*unsurprisingly*".

P15 ll. 302: "*at lower levels"* → change to: at all levels

P15 ll. 305: What do you mean by "*changes*"? Differences between the minimum in May and maximum in Aug.? Please rephrase.

P15 ll. 305: "*Changes in water vapor at 147 hPa (Fig. 6g) are larger in CAMS and JRA-3Q*" → Looking at Fig.6g, I guess you mean MERRA-2 instead of JRA-3Q? Please clarify and revise the full sentence.

In the following up to ll. 309 (P16) it is difficult the follow the description, in particular what figure (panel) you are referring to. I therefore cannot trace the stated biases (5/10 ppm) in ERA5/MERRA-2 given in ll.308.

P16 ll.309: "early part" →maybe change to: In May

P16 ll.331-332: “good qualitative agreement” followed by “persistently high distortions” sounds a little contradictory when written in the same sentence.

P16 ll. 332: “*largest ozone concentrations*” → largest ozone concentrations in Aura MLS and the reanalyses

P19 Figure 9 (In general comments): The use of CERES EBAF data (Fig. 9) appears somewhat out of nowhere here (also in sect. 2.2, P5 ll. 145). Although this figure is interesting and supportive for your conclusions (regarding the overestimation of deep convection by MERRA-2), you can consider whether it is sufficient to append it to the supplement. Here, the OLR difference (CERES-each model) would also be interesting here to demonstrate also differences in OLR between CERES and the other models.

P24 ll. 458 and following: description of Figure 13 is not quite clear to me and could be improved. I would recommend to rephrase it.

P25 ll. 470-471: *“Remarkably, assimilation in ERA5 acts to reinforce rather than reduce high biases in ozone relative to Aura MLS (see Fig. 4, Fig. 6, and Fig. S4 in the online supplement)”*. This result is interesting indeed. However, I see this reinforcement only in Fig. 6, but according to Fig.4 (e) the increments seem to act in the right way (at least positive). Can you comment on this? Maybe P. 26 ll. 509 should be also revised then, in case you decide to change text on P25.

P25 ll. 481: *“Reanalysis water vapor, ozone, and CO products generally compare well with Aura MLS observations, especially in their representations of regional anomalies specific to the monsoon”*

I think this sentence could be improved to do justice to the results of the paper. I agree that the basic regional trace gas distributions are reproduced quite well, but you also found that there are also major differences especially in the seasonal variability between the different models and between model and MLS.

The sentence after ll. 482-483 reads a bit confusing. More explanations are necessary why this is a surprising/unsurprising result (maybe by means of literature references).

P25 ll. 485. Do you mean the moist bias in the lower stratosphere, or which altitude region? I would also recommend to replace the reference Krüger et al., (2022) by e.g., Davis et al. 2017 who show indications for a moist bias in the LS in different reanalyses.

P26 ll.503: “*Data assimilation exerts an even greater control on ozon*” control → influence

P26 ll. 514-515: “*However, assimilation of water vapor retrievals in the stratosphere is not the only route to improving stratospheric water vapor”.*

I fully agree that assimilation of humidity is not the only key to improve water vapor in the stratosphere. Data assimilation is not a tool to remove systematic features – but by nature a data assimilation ideally should act to reduce them. You can also use data assimilation diagnostics to learn a lot about the mechanisms (e.g., dynamics) causing systematic biases.

In P26 ll. 529-531 you provide strong conclusions/recommendations how the assimilation of observations in the stratosphere should be treated in the models.

In my opinion, this recommendation is probably too strong, as it has not been proven in your study whether Aura MLS is exclusively responsible for the assimilation increments.

This could be analyzed by means of an Observing System Experiment (data denial) which allows to attribute increments to the assimilation of particular observation types (e.g., Aura MLS). This is certainly beyond the scope of your work, but maybe you should consider narrowing your statement. Can you comment on this?