

Review of “*Evaluation of stratospheric transport in three generations of Chemistry-Climate Models*” by M. Abalos et al.

This paper summarizes and discusses the results of a very comprehensive comparative evaluation of stratospheric transport in three generations of chemistry climate models that participated in successive model intercomparison initiatives over the last 15 years, including the latest one, CCMI-2022. The authors analyze several direct and indirect metrics of transport, including but not limited to the advective component of the BDC, several mixing diagnostics, wave activity, and polar vortex breakup dates. They analyze the seasonality as well as long-term trends of these metrics against satellite and reanalysis-derived data. One disappointing conclusion is that well-known biases continue to haunt generations of CCMs, and in some cases, such as the age of air, the latest models perform worse than their predecessors. This is an important result pointing to something that we as a community need to tackle.

I don't often say this in a referee report but this is an excellent and important paper. It is very well written, comprehensive, and logically organized. On the first read on several occasions I found myself thinking “this result merits more discussion” only to find more discussion in the next paragraph. As a reviewer it is my job to find problems with the paper but I came up with almost nothing;) I only have a few very minor suggestions, mostly editorial. I recommend that this work be promptly published after a few technical corrections.

Congratulations to the Authors!
Kris Wargan

Specific comments

LL27-28 This is something that people often say but it is not quite correct. As Brasseur and Solomon (2005) explain in Section 5.2.3:

It is sometimes stated that ozone is produced where its mixing ratio maximizes in the upper stratosphere of the tropics and transported downward and poleward to higher latitudes, thus producing the observed structure in the ozone column. But closer examination reveals a different picture. Figure 5.3 shows that the lifetime of odd oxygen is less than a month above about 30 km in the sunlit atmosphere, decreasing to only about a day near 40 km. The tropical ozone mixing ratio maximum occurs between about 30 and 40 km (see Figure 5.10), in the heart of this region of short lifetime. Under such conditions, ozone is essentially in photochemical equilibrium, with rapid production being balanced by rapid destruction — it does not live long enough to be transported away from this photochemically-controlled zone. Since the time scale for meridional overturning of the stratosphere is slow (order of years) compared to these chemical phenomena, the tropics cannot serve as the source region for the large amounts of total ozone found in

The primary region that supplies ozone to the extratropical lower stratosphere is a narrow “transition” zone as shown in Figure 5.11. Air flowing across this zone moves through rapidly enough to retain relatively high mixing ratios compared to those found at lower levels. The transition zone slopes upward in the winter hemisphere, due to latitude-dependent changes in the propagation of the solar radiation that drives photochemistry and determines the lifetime. Thus the source for the extratropical ozone column lies not in the tropics, but in the meridional flow poleward and downward from this transition zone — a zone of competition between transport and chemistry.

L72. I suggest moving the parenthetical clause to right after “CCMs”, i.e. “...the last three generations of CCMs (CCMVal-2, CCMI-1, CCMI-2022) in representing...”

L121 and L124: Which is it, 68.1 hPa or 100 hPa? According to the MLS Data Quality Document it is 100 hPa in version 5.

L192. Missing citation?

L194. CO₂ → CO₂.

L262. Does this mean that you do subsample the model output for comparisons with ACE-FTS?

L267. I suggest stating here what metric is used for wave activity (45°–75° eddy heat flux at 100 hPa).

L280-281. Is it worth noting that this shift toward younger AoA was also reported and investigated in detail in the latest versions of the GEOS model (Orbe et al., 2025)?

L284. “gin” → “in” ;)

LL337–338. Not across **all** metrics, right? Figure 1 and discussion shows that CCMI-2022 has the largest spread. Same is seen in Fig. 4.

LL341-365 and Fig. 4. I understand that one advantage of deriving the overturning mass flux in this way is that this diagnostic can be computed from satellite data. However, as you say, it comes with some serious caveats. It’s not clear to me how useful this discussion is in the context of this study. I suggest either dropping it or making a clearer case for keeping it in.

L383. Why not make 5f a separate figure (a new Figure 10)?

LL481-484. Nice! I really appreciate your considering these possible sources of the discrepancy.

LL495-497. Since you're not defining EHF and EP flux here I suggest citing some generic literature or, alternatively, adding the definition of eddy heat flux ($\overline{V'T'}$) to Table 4, although that would require defining the primes and overbar.

LL525-526. I don't understand how weak downwelling contributes to the young bias in AoA. As it's written my first reaction is that it should be the opposite: slow BDC in the polar regions would lead to older air.

LL536-538 What are "common models"?

Fig. 12. I suggest adding one sentence in the caption explaining that in the NH/SH large positive/negative EHF means more wave activity.

L684. "temperature ... is cooling". That's a pet-peeve of mine: temperature can't be cool or warm. It can be low or high. I suggest rephrasing.

L709. Typo: drop "j".

L750. Again, some metrics show the largest spread in CCM1-2022.

L815-816. I suggest adding two citations here: APARC 2025 for Hunga and Solomon et al., (2023) for wildfires. The first one, released in December, can serve as a canonical reference for the impacts of the eruption, the latter explicitly talks about the increasing importance of smoke injection from wildfires under climate change.

Tables 1–3. Would it be possible to include the number of simulations for each model?

Figures. Sometimes it's difficult to see the yellow line (CCMVal2 MMed) against the shading. See Figs 3, 4, 6, 9, 10, 18a, 20, 19a-c. Other similar figures look fine, e.g., Fig. 11.

Appendix. I don't think it matters at this stage but Tables A1 and 2 are before the list of references and Table A3 is below the references section.

References

APARC, The Hunga Volcanic Eruption Atmospheric Impacts Report. Yunqian Zhu, Graham Mann, Paul A. Newman, and William Randel (Eds.), APARC Report No. 11, WCRP-10/2025, doi: [10.34734/FZJ-2025-05237](https://doi.org/10.34734/FZJ-2025-05237), available at www.aparc-climate.org/publications/, 2025.

Brasseur, G. P., & Solomon, S. (2005). *Aeronomy of the middle atmosphere* (3rd ed.). Springer.

Orbe, C., Takacs, L. L., El Akkraoui, A., Wargan, K., Molod, A., & Pawson, S. (2025). Changes in stratospheric climate and age-of-air in recent GEOS systems since MERRA-2. *Journal of Advances in Modeling Earth Systems*, 17, e2024MS004442. <https://doi.org/10.1029/2024MS004442>

Solomon, S., Stone, K., Yu, P., Murphy, D. M., Kinnison, D., Ravishankara, A. R., and Wang, P.: Chlorine activation and enhanced ozone depletion induced by wildfire aerosol, *Nature*, 615, 259–264, <https://doi.org/10.1038/s41586-022-05683-0>, 2023.