Review of “Quantifying the effect of mixing on mean Age of Air in CCMVal-2 and CCMI-1 models”

In this paper, the authors present a thorough and well thought out analysis of the contributions of mixing to stratospheric age of air in both the most recent generation of coupled chemistry climate models and the previous generation. This work is an important contribution to understanding model differences in tracer transport and in their representation of the circulation. They find that the residual circulation differences are not the primary driver of the age differences between models—rather the mixing efficiency differences are critical. They make an heroic attempt to understand what leads to mixing efficiency differences in the models, concluding that there are substantial difficulties in this analysis in an intermodel comparison.

This is an excellent, valuable paper, and I enjoyed reading it.

I have some relatively minor points for the authors to address. I would also highly recommend a thorough read-through for grammar, as there are misplaced commas (etc.) throughout.
P1L7-8: AoA is not a measure of the strength of the BDC, or at least it isn’t depending on your preferred definition of the BDC (some would equate BDC with residual circulation). I would recommend removing this and just say how it’s defined.
P3L11: Please include some more of the historical mixing literature. E.g. Newman et al. 1986, the eddy diffusivity literature of Nakamura 1996 and related observational follow-ups by Haynes and Shuckburg (2000) and Allen and Nakamura (2001).
P5L8: “Basically … schemes” I found this confusing.
P7L3: mean age is definitely not the same as the residence time
P7L13: ?
P7L23: transport by the residual circulation, not through the residual circulation
P8: please clarify exactly how you calculate mixing efficiency
P8L30: allows linear separation of …
P9L26-30…/Fig. 3: I’ll send you age from MLS N2O which has been calculated using Andrews (2001) relationship (along with the citations). This is an additional new observational constraint if you would like to include it. Let me know if you need any more information about it.
P10L20: Perhaps also include Kovacs et al. 2017 (http://eprints.whiterose.ac.uk/111819/1/acp-17-883-2017.pdf)
P16: Please clarify what was done for this analysis. If the correlation is really of the climatologies, one interpretation of that correlating the mean January upwelling to the mean January RCTT, in which case this is answering the question of whether the seasonal cycles are correlated. Generally though note that it makes sense for age to be less well correlated than RCTT, since age is independent and RCTT depends on the residual circulation, which is used to calculate the upwelling as well.
P17L10: This discussion of mixing efficiency in terms of AoA-RCTT is confusing. I thought Garny et al. 2014 showed that they are related but not mathematically equivalent, because of the dependency of aging by mixing on the vertical velocity.
Generally when writing about mixing efficiency, it is assumed that models far from the multimodel mean are “worse”. Since we do not know that the multimodel mean is right, I would be careful with this language.
P18L10: awk. Perhaps “mixing leads to different magnitudes in the relative enhancement of AoA” or something like that.
P18L20: “exemplary” not sure what you mean here

5.1: How was \Delta(AoA) calculated? what were the tropical edges? This matters in isentropic coordinates, and I'd expect it to matter even more for pressure coordinates. If this is with 20 degree tropics, I would ask that the calculation be redone for 35 degree or turnaround latitude tropics, to see if that makes a difference.
P21L25: I think you mean \Delta(AoA)
P22L13-14: So this is not quite true. AoA difference is a biased measure in the lower stratosphere (and as you note above, it's better in isentropic coordinates than in pressure coordinates), but that doesn't mean it's useless even in the light of vertical diffusion being important.
P23L3-4: "A resulting hypothesis…" It's a shame to bury this sentence in the middle of the paragraph. Start a new paragraph, perhaps?
P24L11-13: Awkward phrasing
P25L31: "different advection schemes with this small sample size." would be more clear
P26L12: "improved" should be "is significantly closer to the multimodel mean"
P27L21: Perhaps add a comment here about diagnosing epsilon differently at different levels? e.g. Ray et al. 2010? Not necessary, but potentially interesting, since the neglect of vertical diffusion isn't as problematic higher up.
P27L23: "showed" rather than "could show"
P27L26: "models'" rather than "model's"
P28L5: "as both are driven by wave driving" or some such
P28L16: "a large" rather than "an excessive" Again, we should reserve judgment until we can actually compare
P29L2: "Demonstrated" rather than "presented"
P29L5-7: "… modeled wave type ratio with the mixing efficiency is very low, the difference in models resolved and parameterized waves does not explain the AoA differences…"

Figures: Stylistically, I think it would help a lot if font sizes were consistent and larger
Figure 6 (b) missing a "c" in mixing efficiency. The R values are hard to read.
Figure 7: choose a different color besides green (or make it dot dash or something) to be colorblind friendly