Review of "Sensitivity of Age of Air Trends on the derivation method for non-linear increasing tracers"

In this study, the authors examine how specific choices for method of converting SF6 concentrations to Age of Air (AoA) impact the resulting value and trend. They use output from the EMAC model, which includes both a linear age tracer and a tracer whose boundary condition matches the observed timeseries of SF6. They evaluate two methods for calculating AoA: In one method the width of the age spectrum is varied, along with th e"fraction of the age spectrum to be considered". In the other method, the "ratio of moments" is varied. The different methods lead to different results, and these results may serve to illuminate the consistent apparent positive trend in midlatitude AoA obtained from previous SF6 balloon measurements (Engel et al 2009, 2017).

Overall this is a good study. The experiments seem well thought out and the analysis is thorough. The result that the trend in age could be this biased by the method for calculating it is extremely important. In fact, this result could be highlighted a bit more. I believe that this will be suitable for publication in ACP following major revisions to improve the introduction and the theory sections, in particular. With these improvements in explanation of context and methods, I believe the manuscript will be much more impactful, and I am excited to cite it in my own work.
See below.

There are a few major points where significant additional discussion would be helpful.

1) A more extended discussion of the methods is needed.
    a. The ratio of moments, for example, is not defined mathematically or explained physically
    b. Although the "fraction of input" is defined, it's not clear what it means. Having never done the calculation this way myself (like the majority of your readers, probably), my intuition for this particular quantity is lacking.
    c. Appropriate citations for 2.1 and 2.1.1 should be included. Hall and Plumb 1994 or Waugh and Hall 2002?. Equation 6 and the discussion around it should be moved to 2.1.1 and this discussion expanded to include the possibility of other shapes such as two peaks (e.g. Andrews et al. 2001). Then the Li et al and Ploeger and Birner references make sense, as they are expanding the discussion by referring to time-varying age spectra.
    d. Why does your convolution method not involve iteration? E.g. Stiller et al. 2012 section 3.3. They find no major difference between using w=0.7 and w=1.5, and this may be due to their iterative process. Since this is one of the main methods that has been used to calculate age from SF6 observations, it should be included here.

       A diagram (or multiple diagrams) may be helpful in expanding this theory section. Show what the difference between w=0.7 and w=1.5 looks like, show what is meant by "fraction of input", etc. Since the crux of this paper is modifying these variables, it seems essential that the reader have a clear understanding of what they are.

2) Some discussion of why the larger ratio of moments causes a lower (more negative) trend seems important. This should be demonstrable analytically, I think. But at least describe the qualitative argument for why this should be the sense of the difference. Adding this will avoid any concern about these results being model dependent or method dependent.
3) The context/introduction for this paper is missing some literature and is a bit incorrect in a few places.
    a. Paragraph 1: Hall and Plumb (1994) does not actually discuss age in quantifying the BDC except with respect to being able to predict transport of other tracers. No vertical velocities appear in that paper. The quantitative relationships between AoA and the circulation were really first established in pressure coordinates using the tropical leaky pipe model in Neu and Plumb (1999). Ray et al. (2016) improved upon this framework, incorporating other tracer information as well. Linz et al. (2016) also built upon Neu and Plumb and isolated the diabatic circulation as a function of the horizontal difference in age of air between the tropics and the extratropics. In a follow up paper, they quantitatively calculated the overturning strength from tracer observations to compare with a model (Linz et al. 2017).
    b. Paragraph 2: There is not really a discrepancy between the apparent increasing trend in age of air observations and an increasing trend in the circulation strength itself. There are a number of reasons for this:
        i. The age of air at one latitude and height is not directly related to the BDC. It is a function of the overturning and of the mixing (e.g., Garny et al. 2014). It is the difference in age between the tropics and extratropics that is quantitatively related to the BDC, specifically the diabatic circulation (Linz et al. 2016, Li et al. 2018).
        ii. Meanwhile, model trend calculations have all been done for the residual circulation in pressure coordinates (e.g. Butchart et al. 2010). The atmospheric circulation itself is expanding (Singh and O'Gorman 2012), which has the result of "accelerating" the stratospheric circulation if one only looks in pressure coordinates (Oberlander-Hayn et al. 2016).
        iii. The observed trend in age found from the MIPAS data in Haenel et al. 2015 is not uniformly positive. Also, that length of record (10 years) is insufficient to determine a real trend in the BDC, due to the internal variability (Hardiman et al. 2017).
        iv. The work by Garcia et al. (2011) demonstrated that with the sparse sampling of the balloon measurements in Engel et al. 2009, a positive trend could easily be found in an overall decreasing timeseries (Fig. 12).

            Although it hasn't been done, between Garcia's results and Hardiman's results, I suspect the additional AirCore points of Engel et al. 2017 would still not be sufficient to properly sample the forced trend.

    c. Paragraph 3: Specify that you are looking at SF6 measurements. There are plenty of other issues that apply to determining age from other tracers, and this study has enough in it without getting into CO2 or non-clock tracers. You probably want to change the title of the paper to reflect this. Perhaps title something like "Are

positive trends in stratospheric age of air from SF6 measurements an artifact of the nonlinearity of the SF6 timeseries?" Hall and Waugh 1998 should be mentioned here also for context.

d. Paragraph 4: Missing relevant literature on SF6 lifetime: associative electron attachment has newly been shown to be dominant beneath 105 km (Totterdill et al. 2015, Kovacs et al. 2017)

Detailed comments:
p.2
l.13 cause→case
l.20 more closely
p.3 l.7 and many other places "allows to": allow takes an object before its verb. So "allows *somebody* to *do something*" is the proper construction. Alternatively, "allows use of… and thereby better understanding of" would work here, since "use" and "understanding" are both objects.
p. 5 l 31 mention here that your SF6 doesn't have a mesospheric sink, so that aspect of uncertainty in AoA calculation is not included in your examination. (a strength of this study, I believe)
p. 6 l 4 Completely interpolated from what?
Fig. 3 and discussion thereof. Why exactly is the resolution only 3 months? I assume this is related to your pulse setup, so perhaps explain it explicitly there.

p.8 l.3 Can you be slightly more explicit about this extension—specifically, how is "second half" defined?
p.8 l.6—does this mean that the inverse Gaussian form is problematic? If so, state that.

Discussion of Figure 6 is confusing. State explicitly what was done, why, and what the results show.

p. 11 l.2 Could this be because of your 9.5 year tracer reset? Unless this actually lines up with the solar cycle well, in which case, show that.

Fig. 8 caption: just repeat the description here for the middle and right panels rather than making the reader go back to Fig. 6.

Discussion of Figs. 10 and 11: I got lost here. Could you add a summary the conclusions from these two after describing them?

p. 19 l. 5 "the trend" what trend? Also, this is not a good topic sentence for the rest of the paragraph. You don't need to talk about methane oxidation, unless there's a specific point to make here.

p. 19 l. 12 Linz et al. 2017 did this, i.e. compared MIPAS SF6 to model(s?) while accounting for the sink, and the sink was very limiting.

p. 20. End with a stronger statement than this! Your work is much more important than your concluding paragraph implies. Something about needing to understand much more about details of age spectra before calculating forced trends? Your study is another excellent example of why there is not actually a clear disagreement between measurements and models—the measurements are both too limited and too limiting to calculate a trend the same way that is possible for the models.

Figures:
Single panel figures have large enough labels, but the labels on multipanel figures are very small. Fig. 7 also has small labels. Label locations for colorbars are not consistent.

Fig. 3: choose another color (or line style) for one of MAM or JJA, as this is not currently colorblind friendly. Could just switch Hall and Plumb color with MAM.

Fig 10 Just one power of 10, probably 14 for both panels

References:
Neu and Plumb, 1999 JGR "Age of air in a "leaky pipe" model of stratospheric transport"
Ray et al. 2016 doi: 10.1002/2015JD024447
Linz et al. 2016 doi:10.1175/JAS-D-16-0125.1
Linz et al. 2017 doi: 10.1038/NGEO3013
Li et al. 2018 doi: 10.1002/2017JD027562
Butchart et al. 2010 doi: 10.1175/2010JCLI3404.1
Singh and O'Gorman 2012 doi: 10.1175/jcli-d-11-00699.1
Oberlander-Hayn et al. 2016 doi: 10.1002/2015GL067545
Hardiman et al. 2017 doi: 10.1002/2017GL072706
Hall and Waugh 1998 JGR "Influence of nonlocal chemistry on tracer distributions: Inferring the mean age of air from SF6"
Totterdill et al. 2015 doi: 10.1021/jp5123344
Kovacs et al. 2017 doi: 10.5194/acp-17-883-2017