Response to Reviewers – "Age of air from ACE-FTS measurements of sulfur hexafluoride" by Laura Saunders et al.

We thank the reviewers for their time and insight, and we appreciate their thoughtful comments. The reviewers' comments are given below **in bold** with our responses in red.

Reviewer 1

This paper presents a new dataset of stratospheric age-of-air obtained from ACE-FTS satellite measurements of SF6. The dataset spans the period from 2004 to 2021, making it the longest satellite record of the age of air. The raw SF6 measurements and methodology are described in detail, and a comparison with other satellite measurements is provided.

In my opinion, the approach is sound, the paper is well-written, and the figures are clear and well-designed. I recommend publication pending a few minor revisions described below.

My two concerns are as follows:

1. **There is limited discussion of the uncertainty in the final age product. The SF6 measurements are presented with error bars representing the sampling uncertainty from ACE, but these are not extended to the AoA product. Additionally, a discussion of estimated uncertainty arising from the SF6 retrieval process and its impact on AoA is missing.**

The AoA uncertainties are shown in Figures 9 and 10, however there were typos in the captions. The last sentence in the caption in Figure 9 used to read: "Error bars correspond to one standard deviation." It now reads "Error bars show the standard deviation of the age in each bin." The last sentence in the caption in Figure 10 used to read: "Error bars correspond to one standard deviation." It now reads "Error bars correspond to one standard deviation of the mean ages from the bins that were averaged together."

In addition, an explanation has been added of how the uncertainty (i.e., the "standard deviation" and "standard error of the mean") on the age of air in each bin is estimated. The following text was added to the end of Section 4.3:

"The uncertainty on the age of air in each bin was estimated in two ways. The first was by propagating the standard deviation of the SF_6 VMR through a simplified age of air calculation. That is, the upper bound on the age was taken to be the lag time between the average SF_6 VMR plus the standard deviation and the equivalent point on the tropospheric reference curve. The lower bound was estimated the same way but using the SF_6 VMR minus the standard deviation. The "standard deviation of the age of air" was taken to be half of the range between these two bounds. The same calculation was also done using the standard error of the mean SF_6 VMR to determine the "standard error of the mean age of air". Both of these uncertainties are provided in the data files."

The following bolded sentence was also inserted into the last paragraph of Section 4.4:

"For each apparent age value at time tmeas,i, that is, the age derived for each bin in the ACE-FTS and MIPAS SF_6 climatologies, Ft(tmeas) was determined for the relevant year and Equation 4 was solved numerically to find the ideal age. **The uncertainty on sink-corrected age of air was estimated by applying the sink correction scheme to the age plus the estimated uncertainty and to the age minus the estimated uncertainty, and using half the range of these two values. This was done for both the standard deviation and the standard error of the mean.** It is important to note that the sink correction only depends explicitly on time and not on altitude or latitude."

On Lines 388-389, the following clarification in bold has been added: "and each bin is shown as one point with error bars representing one standard deviation **of the age of air (as described above)**".

Line 408: The following sentence in bold was added: "For these comparisons, the sink-corrected ACE-FTS ages are shown on the same plots as the uncorrected ages, but as filled markers rather than open markers. **The error bars represent the standard deviation of the mean age of air in the bins that were averaged together across all years.**"

As for retrieval errors, the random error is greatly reduced by taking zonal averages. The statistical uncertainties addressed above tend to be much larger. In addition, a full error budget including systematic uncertainties is not currently available for ACE-FTS retrievals.

2. **A potential weakness in the analysis, particularly in extracting different signals of variability from the time series in Figures 13, 14, 16, and 17, may stem from the limited sampling by ACE of the time-evolving SF6 field. This field is subject to synoptic variability associated with atmospheric motions (e.g., isopleths deformed by Rossby waves). For MIPAS, the redundancy of daily measurements mitigates this effect, but this is not the case for ACE. It seems straightforward to address this issue by using equivalent latitude (and potential temperature) (see, e.g., Allen and Nakamura, 2003; Hegglin et al., 2006).**

In order to address this comment, the analysis was repeated using equivalent latitude bins. The following bolded note has been added to the text at the beginning of Section 4: "In addition, a dataset with 5° latitude bins and one 2 km altitude bin from 19-21 km was also created for comparison purposes. **There is also an age of air product provided in 10° equivalent latitude bins, based on the equivalent latitude provided as part of the JPL DMPs.**"

It was found that there are only minor differences in the zonal mean age of air in most regions, shown in Figure R1. Somewhat more significant differences only occur in the NH polar vortex, which appears better confined (with older air) in equivalent latitude space. Calculating trends in equivalent latitude bins had not been done, as it would not accomplish the goal of determining how age of air is changing in geographical space.

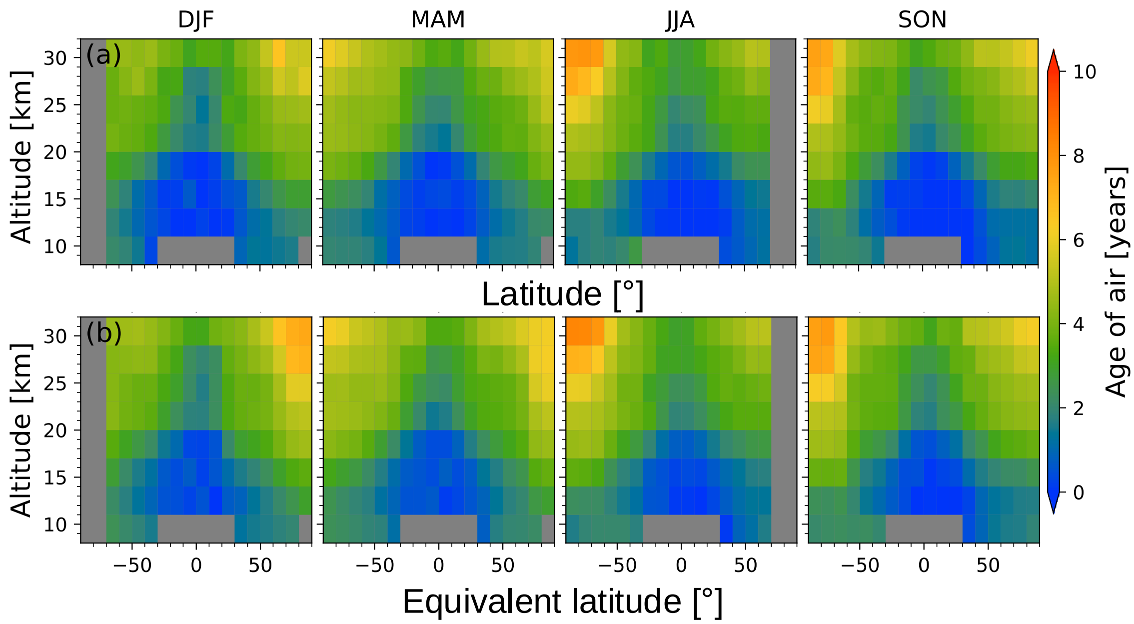

Figure R1: (a) Multi-year zonal mean age of air from the full ACE-FTS v3.5/3.6 dataset (February 2004-February 2021) with latitude on x axis. Each column represents one season. (b) Same as (a) but with equivalent latitude on the x axis.

The analysis was also repeated in geographical latitude on potential temperature levels spaced by 50 K (~2 km). Trends were calculated at the 400 K, 450 K, and 500 K levels, which are approximately equivalent to the 14-20 km region. Here, the trends were still found to be negative and significant in 14 out of 18 bins, and insignificant in the remaining 4 bins. The mean trend found in each hemisphere (NH: -0.11 +/- 0.02 yr/decade, SH: -0.11 +/- 0.03 yr/decade) is consistent with the mean trends in the altitude bins (NH: -0.10 +/- 0.01 yr/decade, SH: -0.11 +/- 0.02 yr/decade). As such, it was decided to keep the analysis as is, in altitude bins.

Other minor comments/typos:

Line 1: "increased" → "increasing"

This sentence describes greenhouse gas concentrations that have already increased. The warming and cooling described later in the sentence are a result of the higher concentrations, not the rate of change of these concentrations.

Line 32: "faster transit times" → "shorter transit times"

This change has been made. In addition, "slower transit times" has been changed to "longer transit times" at the end of the sentence.

Lines 32-34: The sentence "Since neither models nor observations allow the examination of infinitesimally small air parcels, it is necessary to address the fact that any region under consideration is made up of air of different ages." is confusing and may be deleted altogether (see comments below regarding infinitesimally small air parcels).

This text now states: “It is also necessary to address the fact that any finite region under consideration (that is, a macroscopic “air parcel”) is made up of air of different ages.” This is included to clarify what is meant by “The distribution of these ages” in the sentence that follows.

Line 39: Consider changing "calculated" to "estimated."

This change has been made.

Line 105: "high-resolution" → "high spectral resolution"

This change has been made.

Lines 120-123: It might be useful to provide more information regarding the sampling by ACE.

This change has been made as follows.

The text previously read: “Due to the nature of solar occultation measurements, ACE-FTS has limited sampling, which must be taken into consideration when comparing ACE-FTS with other instruments or with models,”

The text now reads: “Due to the nature of solar occultation measurements, ACE-FTS has limited sampling. It takes three months for ACE-FTS to obtain global coverage and the measurements are not distributed evenly in latitude throughout each month. This must be taken consideration when comparing ACE-FTS with other instruments or models,”

Line 150: Have the authors tried different distances in the longitudinal and latitudinal directions? Using equivalent latitude may be an option here.

Using latitude and longitude for the spatial coincidence criteria is an option, but it does not change the ACE-MIPAS comparisons nor the number of coincidences in a meaningful way.

This is demonstrated below by showing comparisons with two different coincidence criteria in Figures R2 (1000 km, same as Figure 1) and R3 (300 km). The comparisons are nearly identical because the majority of coincident MIPAS measurements are within 200 km of the corresponding ACE-FTS measurement (equivalent to a maximum distance of about 2° latitude). This is made possible by the relatively generous 12 h time criterion that was used, which is reasonable since SF_6 does not have a diurnal cycle.

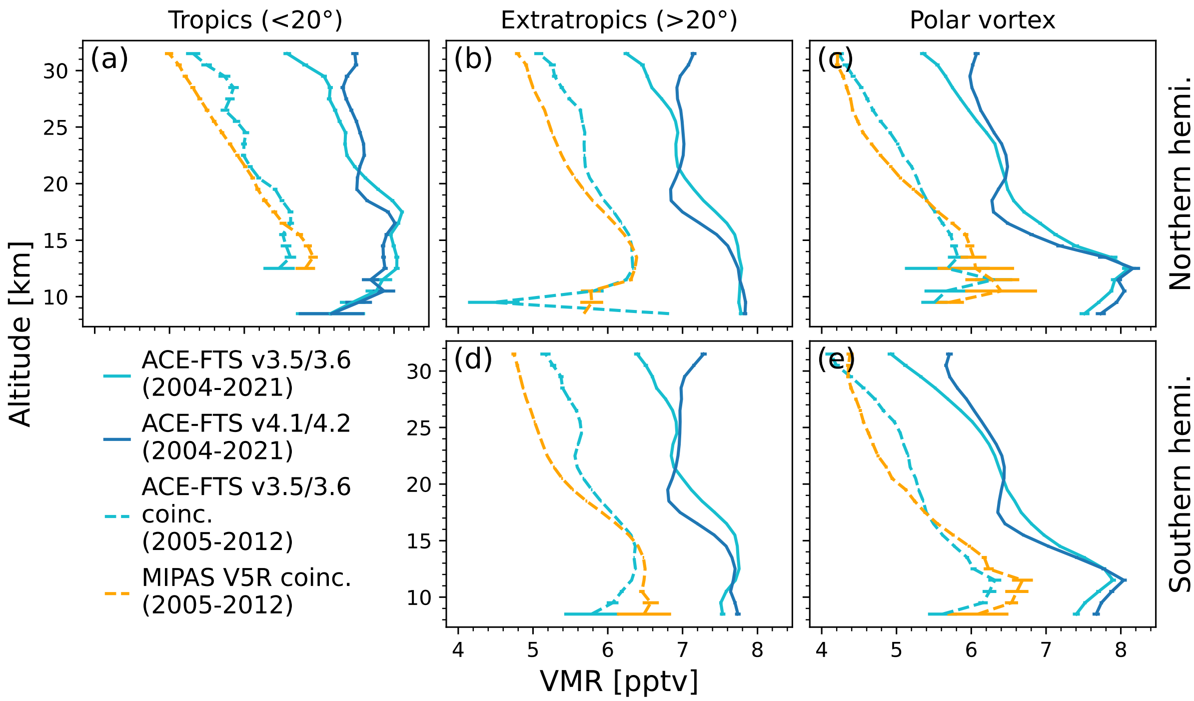

Figure R2: Same as Figure 1 in the manuscript (1000 km coincidence criterion).

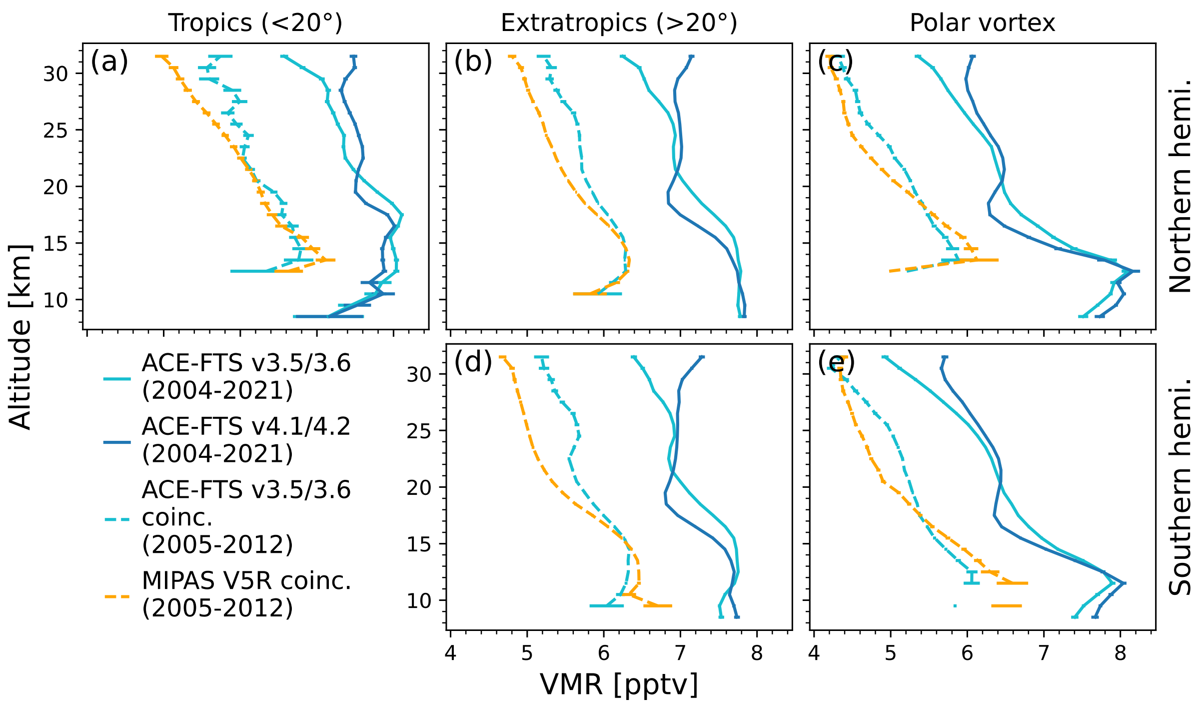

Figure R3: Same as Figure 1 in the manuscript, but with a 300km coincidence criterion used for distance.

Using equivalent latitude is a useful way to address differences in concentrations due to the polar vortex, but the need for this was mitigated by using the polar vortex edge coincidence criterion.

Line 166: "one standard error" → "one standard deviation"

Here, the standard error of the mean was used because the comparison is between two mean profiles. In other instances where a mean profile was compared to individual profiles, the standard deviation was used.

Line 173: Change "Fig. 1" to "Figure 1," as per ACP convention (it is the subject of the sentence).

This change has been made.

Figure 1: It might be worth showing the average for the entire MIPAS dataset to get a sense of the uncertainty caused by ACE sampling.

This analysis was done and is shown in Figure R4 below. The mean from the full MIPAS dataset is not too distinct from the sampled dataset and its inclusion makes it more difficult to see the comparison between the coincident datasets.

The following bolded sentence was added: "This reinforces the importance of using coincident measurements when comparing observations from two different instruments; comparing the full ACE-FTS mission with the full MIPAS mission would introduce significant bias due to the two different time periods. **The mean profile from the full MIPAS mission is similar to that from the sampled MIPAS dataset, and is not shown because it makes the comparison between the mean coincident profiles more difficult to visualize.**"

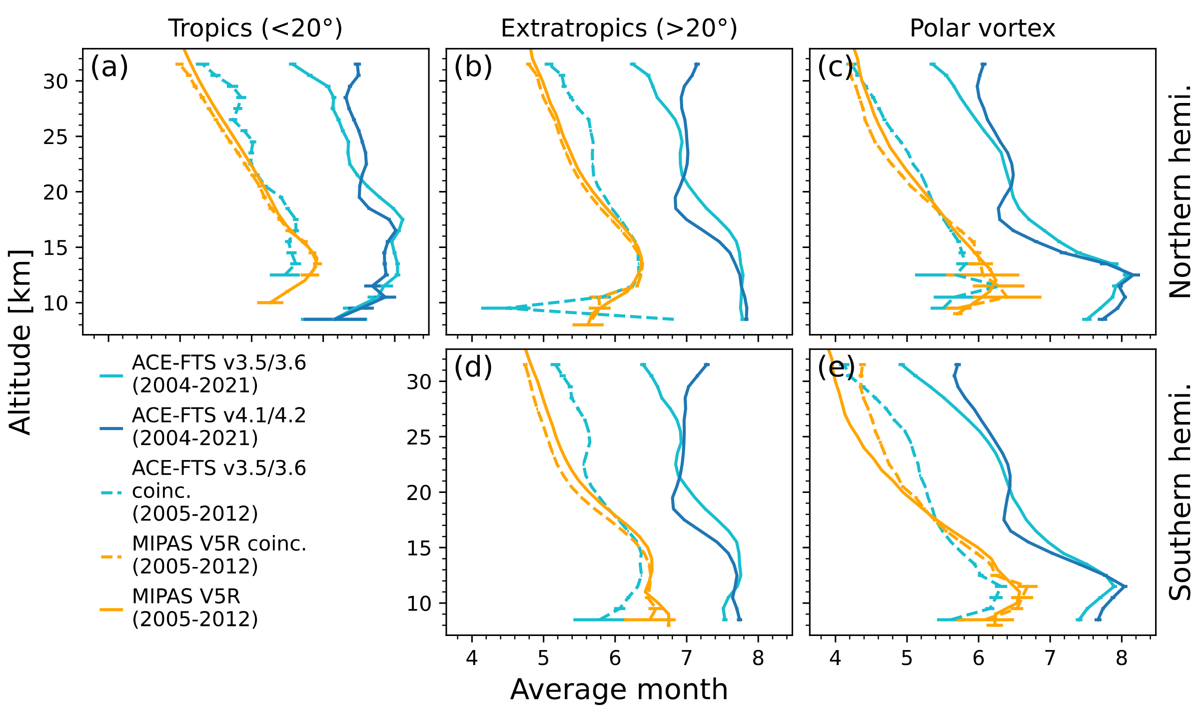

Figure R4: Same as Figure 1 in the manuscript but with the mean of the full MIPAS V5R dataset (solid orange line) included.

Line 261: Consider removing "...were actually infinitesimally small and therefore...". In practice, it is only at the molecular/free path scale that mixing/diffusion, and thus the mixing ratio, does not make sense.

This change has been made.

Line 267: Consider using another symbol for the ratio of moments instead of w.

This symbol was used to follow the formalism used in i.e. Stiller et al. 2012.

Line 267: Add "years" to "constant value of 0.7".

This change has been made.

Line 282: Shouldn't the left-hand side of Equation 2 read [SF6]modeled? Further detail on the numerical evaluation of the integral should be provided, such as the discretization step along transit time and the actual upper boundary of the integral (which I imagine is finite).

The left-hand side does refer to the measured SF_6 value, since this is what is being used to determine the mean age. The descriptions of the discretization step and upper boundary of the integral have been added to the text (in bold): "This process is repeated, **using $4\Gamma_{new}$ as the upper bound on the integral with 1000 intervals**, until the difference between $[SF_6]_{modelled}$ and $[SF_6]_{measured}$ is below 0.01 pptv, which is well below the standard error of the mean SF_6 value in each bin and typically allows convergence within five iterations."

Line 292: As far as I understand, this is Newton's method. You might mention this explicitly.

The text previously read: "A new guess for $t_{\_trop}$ is then calculated by moving along the reference time series by the difference in SF6 values:"

The text now reads: "A new guess for the tropospheric time ($t_{\_trop,\ new}$) is then calculated using Newton's method:"

Line 375: Which tracers would not be affected by this difference in vertical gradients?

This is specific to SF_6 and the sentence has been clarified. The bolded text has been added: "Future studies should include comparisons to datasets with ages derived using other clock tracers **where ACE-FTS and MIPAS have better agreement**."

Lines 377-378: I recommend either making the first half of the sentence more quantitative by mentioning the accuracy of the different measurements or removing it.

This has been updated.

The text previously read: "In situ measurements are generally considered to be more precise than remote sensing measurements, so they are a particularly useful reference for comparison.

The text now reads: "In situ measurements are a particularly useful reference for comparison."

Figure 11: Would it be possible to generate a similar wing plot for the SF6 measurements during the MIPAS time period? As it stands, it is not entirely clear where this difference in age stems from.

This has been done and the result is shown in the second panel of Fig. R5 alongside the original figure. The differences between ACE-FTS and MIPAS do not improve significantly by considering only coincident measurements (2005-2012), so the difference is not due to sampling.

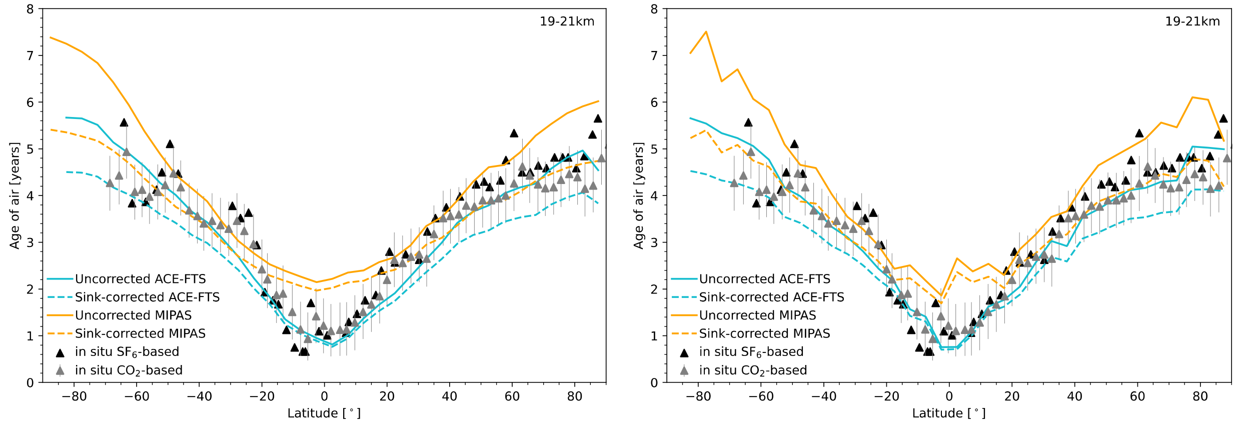

Figure R5: (a) Same as Figure 11 in the manuscript. (b) Same as Figure (a) but the ACE-FTS and MIPAS ages are only based on measurements that are coincident between the two datasets.

Line 540: For the impact of the sink correction on the trend, Loeffel et al. (2022) may be a better reference.

This change has been made.

Line 575: Is this impact of decadal variability in MIPAS also found in ACE data?

Yes, but the full ACE-FTS time series is long enough to see that this is indeed decadal variability and not an overall change in age. This can be seen by comparing Figures 13 and 14 (full time series analysis – 2004-2021) with Figures 16 and 17 (MIPAS time period analysis – 2005-2012). Changing the time period changes the sign of the trend, so yes there is an impact on ACE-FTS data.

Reviewer 2

This manuscript describes a new stratospheric age of air (AoA) dataset derived from ACE-FTS measurements of SF6. Spanning 2004-present, this is the longest continuous AoA dataset available, and it represents a timely and valuable contribution to the ongoing question of whether AoA is changing with time. Overall, I find the manuscript to be well-written and the presentation of the derivation of AoA from SF6 and comparison to satellite and in situ measurements to be robust and informative, though lacking key details regarding sampling density. However, I find that there are significant issues with the analysis and discussion of AoA trends that warrant major revisions prior to publication in ACP.

Firstly, I find the discussion of trends based on comparisons to balloon measurements to be a significant over-reach, even with the few caveats provided about the representativeness of the data. Given the large interannual variability in AoA shown, for example, in the manuscript's own Figures 13 and 14, it is impossible to draw inferences about trends from 6 profiles taken at different times and locations. The balloon comparisons are useful for examining the vertical structure of AoA, but should not be used as a basis for inferring changes in AoA with time.

We believe it is pertinent to point out the consistency of these results with a decrease in age of air over time. We have removed all instances of the words "indicates" or "implies" and instead refer only to consistency. The purpose is not to use these comparisons as evidence of a BDC acceleration. The edited text should ensure that this is clear.

Lines 418-420 previously read: "This could imply a change to the northern deep branch of the BDC between 1995 and the mid-2000s, but could also be due to interannual variability."

They are now Lines 431-432 and read: "This is consistent with a change to the northern deep branch of the BDC between 1995 and the mid-2000s, but could also be due to interannual variability."

Lines 428-430 previously read: "This could also be an indication of BDC acceleration, but the difference is small relative to those seen for SESAME 95 and OMS 1996, and the presence of the polar vortex significantly complicates the comparison."

They are now Lines 441-442 and read: "This is also consistent with a BDC acceleration, but the difference is small relative to those seen for SESAME 95 and OMS 1996, and the presence of the polar vortex significantly complicates the comparison."

Lines 591-593 previously read: "Comparisons with age of air derived from balloon measurements of SF_6 and CO_2 from before the ACE-FTS mission suggested that the shallow branch of the BDC might be accelerating, and possibly that the deep branch might have slowed between 1995 and the mid-2000s,…".

They are now lines 604-606 and read: "Comparisons with age of air derived from balloon measurements of SF_6 and CO_2 from before the ACE- FTS mission are consistent with an acceleration of the shallow branch of the BDC, and possibly a slowing of the deep branch between 1995 and the mid-2000s,…".

Secondly, the analysis of trends using the full ACE-FTS dataset requires 1) either a different approach than using the solar cycle to represent unknown decadal-scale variability or a much more robust justification for doing so, as it seems to actually introduce some decadal-scale variability that may not be present in the data, and 2) the provision of much more information regarding the statistical robustness of the trends shown. Given the current focus on trends in AoA as potential markers of changes in the stratospheric circulation, it is crucial that the trend analysis be carefully and judiciously considered.

The fit components shown in Figure 15 and in the supplementary material demonstrate that there is significant decadal-scale variability in the time series that is well-represented by an offset solar cycle. The fit residuals also make it clear that a decadal-scale cycle (such as the solar cycle) is necessary to produce a reasonable fit. An example is shown in Figure R6 using the 50-60° N bin at 14-17 km. The left side shows the fit with the solar cycle and the right side shows the fit redone without the solar cycle. For the latter, the distribution of the residuals has two modes, and the residuals are generally larger. The fit does not represent the data well and the other proxies are also clearly affected, as shown in Figures R7 and R8, which show the proxy contributions for each of the fits in Figure R6. In particular, the contribution of the ENSO proxy is greatly inflated when the solar cycle proxy is not included. The annual cycle also has a much smaller contribution.

The text now includes a discussion of magnitudes. The previous text on lines 527-536 read: "The linear term in most bins, especially at lower altitudes, was sensitive to the inclusion of the decadal variation in the regression model; without it, the trend could even change sign. This indicates that the variation in age related to the decadal variation is similar to any long-term change seen over this 17-year period and reinforces the importance of having a long time series for the detection of the linear trend in age of air. The ACE-FTS mission currently offers the longest continuous observational time series, but this sensitivity shows that the trend is so small in magnitude that a longer time series would be needed to determine trends with more certainty. Despite this, the trends are consistently negative in the regions being considered here, providing some confidence in the sign. In addition, the negative trends were statistically significant (to two standard deviations) in all bins."

This now reads: "The solar cycle proxy seems to account for this oscillation well; in every bin, the fit residuals are smaller and have less structure when it is included in the fit. However, even if the time series is fit without the solar cycle term in Eq. 5, the linear trend remains qualitatively the same (i.e., negative) and is significantly different from zero to two standard deviations. Therefore the finding that age of air decreases over the full observation period is independent of the assumption of a solar cycle. The mean trend in the midlatitude lower stratosphere is -0.10 ± 0.01 years per decade in the northern hemisphere and -0.11 ± 0.02 years per decade in the southern hemisphere, where the uncertainty is the standard error of the mean."

We feel that this addresses both this comment and the related comments below.

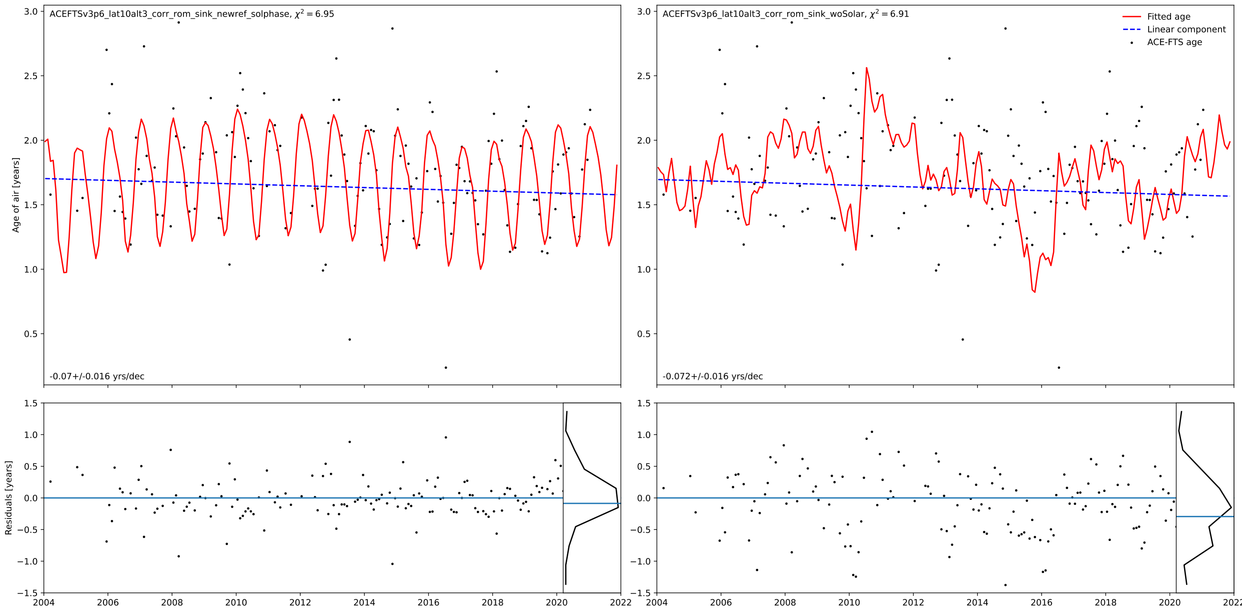

Figure R6: Left: The top panel shows the monthly age of air at 50-60°N, 14-17km with the fit that includes the solar cycle. The linear component is shown with the blue dashed line, and the residuals with their distribution are shown in the bottom panel. Right: the same as the left panels but the fit does not include the solar cycle.

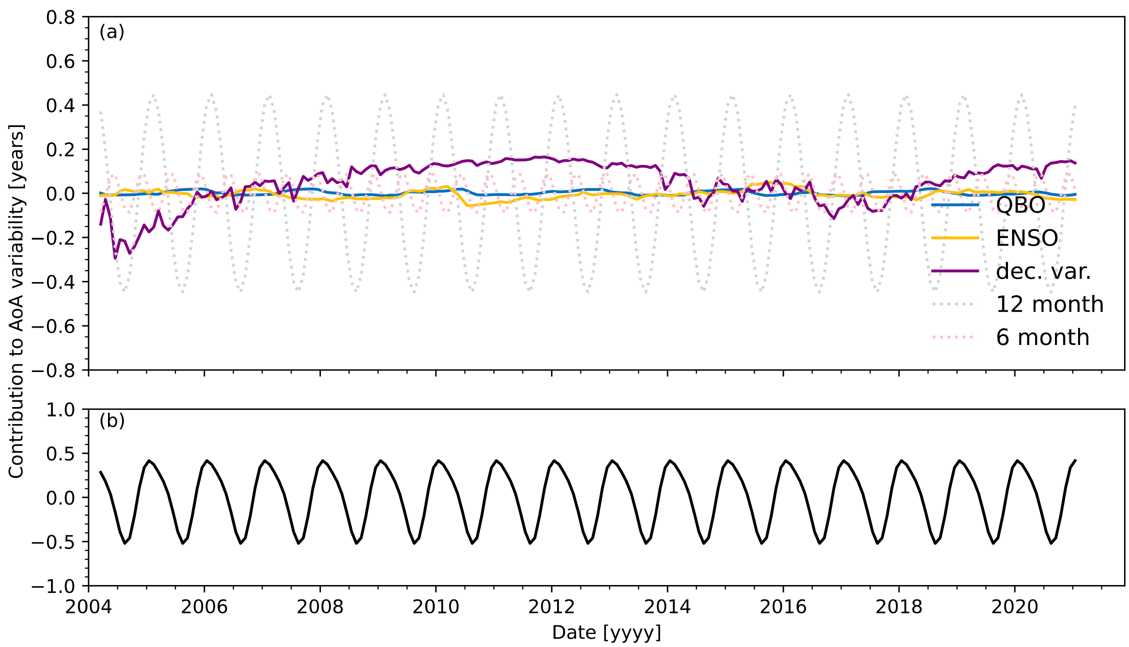

Figure R7 (a) The contribution (in years) of each type of interannual variability to the fitted age of air for 50-60°N, 14-17 km as shown on the left side of Figure R6. (b) The total contribution of the 12- and 6-month seasonal cycles.

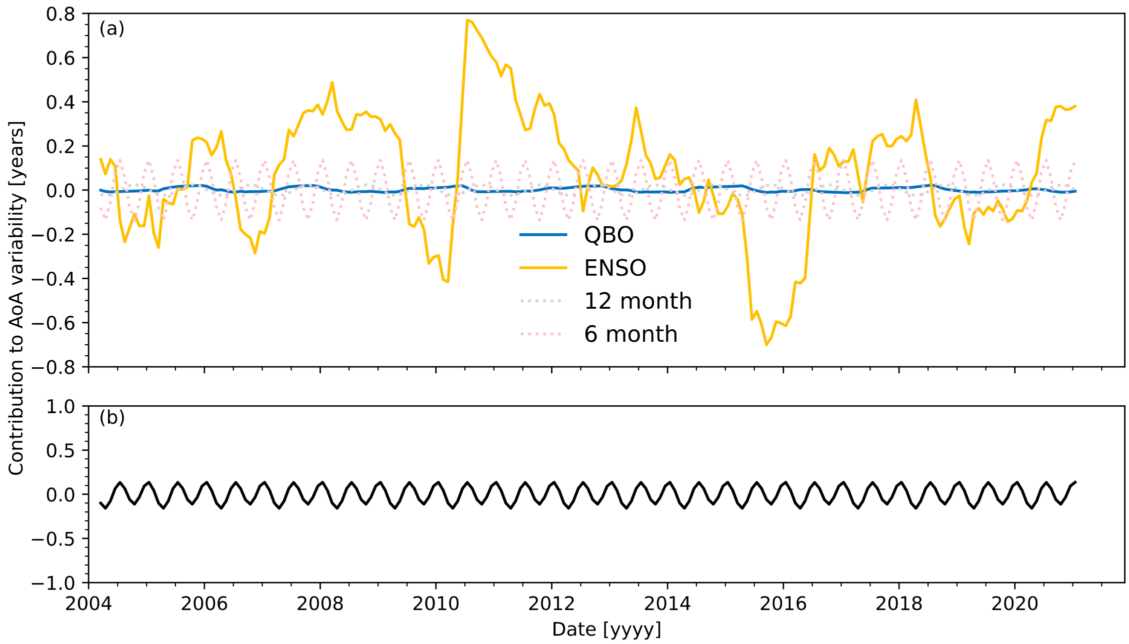

Figure R8 (a) The contribution (in years) of each type of interannual variability to the fitted age of air for 50-60°N, 14-17 km as shown on the right side of Figure R6 (that is, without the solar cycle proxy). (b) The total contribution of the 12- and 6-month seasonal cycles.

My specific comments are as follows:

Lines 25-27: The statement "this has thus far only been indicated by measurements in the northern lower stratosphere" seems to contradict lines 88-90, which say, based on both the Ray et al. study and MIPAS observations, that "the shallow branch in the southern hemisphere does appear to have accelerated."

This has been amended.

This sentence previously read: "Model studies suggest that both branches of the BDC are accelerating (e.g., Butchart, 2014, and references therein), but this has thus far only been indicated by measurements in the northern lower stratosphere (Ray et al., 2014)."

It now reads: "Model studies suggest that both branches of the BDC are accelerating (e.g., Butchart, 2014, and references therein), but this has been difficult to confirm with measurements."

Lines 35-36: "Mixing", which appears here to refer to large-scale mixing of air from different regions (see a comment below on the use of this term) has a one-sided effect on the spectrum (you cannot get younger ages through mixing) and thus leads to a change in the mean age in addition to widening the spectrum.

We use "mixing" here and elsewhere to refer to turbulent-scale mixing between air parcels, which would widen the age spectra on either side since younger air can move into a neighbouring parcel. This has been clarified by modifying the sentence. "Mixing leads to a wider age spectrum […]" has been changed to "Mixing between air parcels leads to wider age spectra […]".

Lines 36-38: The manuscript uses the parenthetical "(with no change in the Tropics)" in several places when discussing how mean age trends in the high latitudes are related to changes in the circulation. The authors need to provide a brief narrative on why this parenthetical is needed (it is because only changes in tropical-extratropical age gradients actually reflect changes in the circulation independent of large-scale mixing changes) and reference the proper papers on this (primarily written by M. Linz).

This is used because the difference in age between the tropics and extratropics is what indicates the speed of the BDC. This has been clarified.

The text previously read: "A decrease in the mean age of air at higher latitudes (with no change in the tropics) indicates an acceleration of the BDC, motivating the need for trend detection in age of air derived from observations."

The text now reads: "A decrease in the mean age of air at higher latitudes, with no change in the tropics, indicates a decrease in the mean transport time from the tropics to higher latitudes, and therefore an acceleration of the BDC (e.g. Neu and Plumb, 1999; Linz et al., 2016; Linz et al., 2017). This motivates the need for trend detection in age of air derived from observations."

Line 39: Please add "monotonically" before "increasing"

Clock tracers do not need to be monotonically increasing. CO_2 for example has a seasonal cycle and, as such, is not monotonically increasing. The seasonal cycle is just a complication that needs to be considered in the calculation of age of air.

Line 142: Consistency with what?

There is a bias between the V5H and V5R datasets, so rather than accounting for this bias just to add a few months of V5H data, it is preferable to use only the V5R data. This has been clarified.

The text previously read: "In this study, the lower resolution SF_6 measurements from version 5 (V5R) (Haenel et al., 2015) were used to maximize temporal coverage while maintaining consistency."

The text now reads: "In this study, the lower resolution SF_6 measurements from version 5 (V5R) (Haenel et al., 2015) were used to maximize temporal coverage while using a single dataset."

Lines 170-171: The decrease in SF_6 below 14 km in ACE-FTS is mentioned several times in the manuscript. The authors should describe the basis on which it is determined to be unrealistic.

The lowest retrieved altitude for ACE-FTS profiles varies over the course of the year, therefore the number of profile points per bin is more variable below 14 km. This is most likely the cause of this unrealistic feature.

The text previously read: "It should be noted that the decrease in SF_6 below about 14 km in most of the ACE-FTS profiles is not realistic."

The text now reads: "It should be noted that the decrease in SF_6 below about 14 km in most of the ACE-FTS profiles is caused by the fact that the altitude of the lowest retrieval varies over the course of the year. Therefore the low-altitude bins can be missing measurements from certain times of year, resulting in this feature."

Line 204: It would be clearer if the heading were "AoA Methodology" or "Methodology for Deriving AoA"

The heading is now "Methodology for deriving age of air".

Line 208: I do not understand what is meant by "separated by year" here – can you please clarify?

This description is to indicate that all Januaries, for example, have not been averaged together. The following bolded parenthetical statement has been added to clarify: "To reduce noise, the satellite instrument datasets were first zonally averaged by month, separately for each year **(i.e., each bin corresponds to one month in one year).**"

Lines 234-239 and 258-259: 2 km below the tropopause probably successfully separates stratospheric and tropospheric air for ACE-FTS, but does it really work for MIPAS with its significantly coarser vertical resolution?

2 km was chosen to accommodate both the 3-4 km resolution of MIPAS in this region and the 3 km resolution from the field of view of ACE-FTS. With these vertical resolutions, a measurement taken 2 km below the tropopause would be sensitive to the region 0-4 km below the tropopause.

Line 261: Here the term "mixing" is used to describe turbulent-scale interactions between air parcels rather than large-scale stirring that would still transport an infinitesimally small parcel to different regions of the stratosphere. Since the word mixing is used to describe what is actually large-scale stirring in other places in the manuscript, the distinction should be made clear here.

As stated above, this is what was meant by "mixing" above as well (see lines 35-36). No changes were made here.

Line 272: For clarity "than ERA-Interim" should be added after "for ERA-5"

This change has been made.

Figure 4 and all latitude-height cross sections: For the purpose of understanding the robustness of the results, companion figures should be provided in the Supplementary materials that provide the number of observations in each bin.

These figures are now provided in the supplementary materials.

Line 416: The phrase "Aside from this difference" does not make sense here because a difference was not actually described in the previous sentence. The sentence beginning on Line 416 stands alone without that phrase.

The bolded sentence has been added to clarify: "A common feature in the in situ ages, visible in the profiles from AIRE 93 (September 1993), SESAME 95 (March 1995), and OMS 1996 (September 1996), is a constant age above about 20-25 km. **The ACE-FTS profiles do not have this feature.** Aside from this difference, the ACE-FTS ages agree with the AIRE 93 (September 1993) ages below 30 km."

Line 417: Isn't it self-evident that the difference at 30.5 km is related to differences in the vertical gradient?

Yes, this statement has been removed.

Lines 418-428: I realize that the authors are using the word "could" here and that they point out here and in lines 429-430 that the variability is large, but I do not think these profiles should be used to say anything about trends. In fact, in the conclusions (lines 591-593), they are considerably more bold, stating "Comparisons with age of air derived from balloon measurements of SF6 and CO2 from before the ACE-FTS mission suggested that the shallow branch of the BDC might be accelerating, and possibly that the deep branch might have slowed between 1995 and the mid-2000s" before reminding the reader that this is based on comparisons with single balloon flights. The variability in AoA is so large (see Figures 13 and 14 of this manuscript) that single-point comparisons like these should not be used even speculatively to infer trends. I recommend that the authors focus on the vertical profiles of age for the balloon-based comparisons.

Please see the response to the first major comment above.

Lines 471-473: Is it not also possible that this could be due to an issue with the vertical age gradients in ACE-FTS AoA seen in Figure 10? Those results compare two uncorrected SF6 AoA datasets and so should not be influenced by problems in accounting for the mesospheric sink.

Yes, this has been dealt with as follows. The text previously read: "This is not necessarily entirely due to an issue in the model; it is possible that the sink correction does not fully account for the high age bias in ages derived from SF_6 measurements."

The text now reads: “This is not necessarily entirely due to an issue in the model; ACE-FTS ages were also shown to be older than in situ ages at higher altitudes in Figure 10. It is also possible that the sink correction does not fully account for the high age bias in ages derived from SF_6 measurements.”

Lines 482-484: Again, the number of measurements in each bin (as a function of time) should be provided in the Supplemental materials.

As noted above, this is now provided in the supplementary material.

Lines 495-498 and 508-514: I simply do not understand the motivation for using the solar cycle as a proxy for decadal-scale variability of undetermined origin. The solar cycle has a very distinctive and asymmetric signal and this methodology imposes that signal without a clear physical reason for doing so. Looking at Figure 13 panels E and F, which are called out in lines 525-526 as being most noticeable with respect to the decadal variability, it is clear that the residuals in the first half of the dataset, when the solar cycle term causes the fit to start low and increase with time, are more positive than the residuals during the second half of the time series when the solar cycle term causes the fitted timeseries to decrease. So I see the decadal variability in the fit but not in the data. Why didn’t the authors remove the other variability and then try to fit a harmonic to the decadal-scale if they wanted to try to account for it. From my perspective, the clear message about the decadal-scale variability is that the trends we quantify from 17 years of data are highly unlikely to be significant, not that we should be using unphysical methods to try to remove it.

Please see above response to the second major comment, which shows an example of the residuals. From this, we see that trying to fit the other variability without including a decadal scale cycle (such as the solar cycle) does not produce reasonable results, so this method would not work. Importantly, the deduced trends remain qualitatively the same independent of including the solar cycle in the fit, as said above.

Line 508: “enso” should be “ENSO”

This change has been made.

Lines 533-536: The actual slopes and uncertainties should be provided either in the figures themselves or in a Table, and the results should be discussed in terms of whether the changes are significant or not.

We have chosen to provide the mean trends in each hemisphere instead of the slopes and uncertainties. The mean value, with the standard error of the mean, provides a more reliable and robust estimate of the trend and shows the significance of the trend in each hemisphere.

Lines 565-566: Again, if these trends are going to be used to infer changes in the circulation then a discussion of the magnitudes and uncertainties is required.

Please see our response to the second major comment and the above comment regarding lines 533-536.