Review of "Climatology and trends in stratospheric mean age of air from correlations of multiple trace gases measured by ACE-FTS" by Voet et al.

The authors apply a multi-tracer method (introduced in Voet et al. 2025) to ACE-FTS v3.6 and v5.2 to derive an improved mean age of air (AoA) estimate. In this method, AoA is estimated from N_2O, CH_4, CFC-11, CFC-12, and HCFC-22 via lookup tables that are calibrated against SF_6-derived AoA. First, the method was tested with CLaMS-based synthetic ACE-FTS measurements. They then compared climatologies with CLaMS/ERA5 and in situ AoA (compiled dataset, aircraft, and AirCore) and derived 2004-2021 trends. Main findings are:

- A roughly 50% reduction in the uncertainty of zonal-mean AoA.
- A claimed around 0.3 years uncertainty for individual profiles.
- Suppression of the ACE-FTS v5.2 SF_6 retrieval artefact.
- A hemispheric dipole in AoA trends, with increasing age in the NH and a decreasing age in the tropics and SH.

The topic suits ACP, and is a successful extension of the previously published method. A less noisy, spatially resolved satellite AoA product would be valuable. Overall, the article is well-structured, although some sections are very long, and readers would benefit from a clearer organization in those parts. In addition, the paper contains a number of inconsistencies between text and figure captions. Below are some points regarding content that the authors should address, as well as technical comments. I recommend **minor revisions**.

Minor comments

Use of equivalent latitude:

Equivalent latitude was used to account for the dynamical characteristics of different air masses. Although it is stated that the equivalent latitude based on EAR5 was used, some questions remain open:

- The specification how equivalent latitude is computed. Assuming it is PV-based equivalent latitude, how is it treated near and below the tropopause and in the tropics where equivalent latitude is poorly defined?
- The five equivalent latitude regions leave a gap between 60°N and 65°N (L209-2011). How are observations in this range are treated?
- The LUTs are built in equivalent latitude. The final products and in situ comparison are binned in geographic latitude (L246-248). Is there a possibility of a discrepancy that could affect the conclusions?

Synthetic proof of concept:

In the synthetic setup (L267-280), the model tracers are first averaged over 15-day x 10° equivalent latitude x height level bis. Gaussian noise, with a standard deviation derived from the measurements, is then added independently for each tracer. As I understand, within a bin, the synthetic values therefore consist of the model bin mean plus a random noise. I am having a little trouble assessing the variability/noise.

- The noise is assumed to be independent across tracers. In reality, the natural variability is part of the empirical standard deviation, which may correlate across tracers. As this is not taking into account, could this overstate the benefit of averaging the five tracers?
- Is the similar spread in Fig. C1 and C2 expected by construction, because the synthetic noise was derived from the empirical standard deviation on the measurements?
- Again, the noise is constructed in equivalent latitude bins, whereas the evaluation uses geographic 10° latitude bins. Please comment on that.

Comparison with in situ data:

- The comparison with in situ AoA mainly tests whether the multi-tracer AoA reproduces precise SF_6-derived AoA. The multi-tracer AoA is a denoised version of the ACE-FTS SF_6-derived AoA, which share the same age definition to the in situ AoA. Discrepancies may arise from the LUT, from the ACE-FTS tracer retrievals, or from differences in the sampled air masses given the large coincidence windows. I believe that the platforms used here (aircraft and AirCore) also measured some of the trace gases (e.g. CFC-11 and CFC-12). Would it make sense to apply the ACE-FTS-based LUT to these in situ observations? This would remove the coincidence error and test the LUT directly. Possible calibration calibration-scale differences between ACE-FTS and in situ tracer measurements should be accounted for.
- When it comes to AirCores, it is important to consider vertical uncertainty (e.g. altitude range or altitude kernel), especially for higher observations. This is particularly relevant when discussing outliers (L475-477). That should be included in the discussion.

Discussion of conflicting recent observational evidence

Ray et al. (2026, Nature Geoscience) report a decrease in observed stratospheric mean age that is consistent with circulation acceleration. This paper was cited in the introduction (L46) but was not discussed further in Sect. 3.4 or 4. The conclusion of slowdown of NH circulation should be put in context with this result as well, including possible reasons for the discrepancy.

Structural and technical comments:

- L116: I'm not sure if the formulas for CFC11, CFC12, and HCFC22 are necessary, especially since they appear in full earlier in the text.
- L130: v5.2 begins in December 2004, while trends are given for 2004-2021. Please clarify.
- L163: Section 2.3 is quite long. It might make sense to divide this section into two separate subsections: the classic SF6-based version and the multi-tracer version.
- L176: “approximations” -> “approximation”.
- L232: conflict in Fig. C1 caption: the text says the example is 30°S-30°N, the caption says 30°N-60°N. The caption also refers to CFC-12 only, although five tracers are shown. Also the time range contradicts itself and the window given in the text.
- L260-161: The appendix is too long. Consider creating a “Supporting Information”.
- L303: “(Sect. 3.1), 2)”; “(Sect. 3.2), 3)”; “(Sect. 3.3), and 4)”. I recommend rewriting this to make it easier to read.
- L314-315: conflict in Fig. 1 caption. The text gives DJF 2011/2012 and JJA 2012, the caption gives Dec 2010-Feb 2011 and JJA 2011.
- L349: conflict on Fig. D2 caption. June-August 2011 vs. 2012.
- Fig. 2 caption: “than” -> “then”.
- L358: “Full” -> “full”; “coverd” -> “covered”.
- Fig.3 caption: ACE-FTS at the end of the first sentence should be removed.
- L396: Section 3: Please consider including the description of the in situ observations (L412-421) in the Methods section of the paper. There, you can also go into a bit more detail about the aircraft and AirCore observations.
- L525: Isn't a period of more than 30 years better for long-term trends?
- L580: Again, I'm not sure if the formulas for CFC11, CFC12, and HCFC22 are necessary.
- L592: “Figures 7 to 6” -> “Figures 6 to 8”.
- L614: “an” -> “and”.
- Fig. A1 caption: the points in panel b are purple squares, not blue dots.
- Fig. A2 caption: this is the exact same caption as in Fig. A1. I assume it’s for v5.2.