Review of "Measurement report: Greenhouse gas profiles and age of air from the 2021 HEMERA-TWIN balloon launch" by Schuck et al.

General Comments:

The manuscript outlines measurements made during the August 2021 large-balloon-assisted launch of the HEMERA-TWIN gondola, which hosted discrete flask sampling systems, AirCore samplers and continuous in situ instrumentation for measurement of long-lived atmospheric trace gases. Profile data are compared to flask sampler data, and stratospheric mean age of air is calculated from flask and AirCore samples of CO2 and SF6.

Overall, this measurement report provides updated large-balloon observations from Kiruna and new and important observation-based calculations of stratospheric mean age of air worthwhile of publication. Some scientific, technical and organizational comments below are recommended to be addressed before publication, however (see specific comments below).

- There are several references to measurements made, but not used in this manuscript and/or described elsewhere (Laube et al., 2024 – not currently accessible). Text throughout the manuscript could be tightened up a bit to not cause undue confusion.
- More details are needed in the onboard sampling systems and analytical descriptions, in addition to the mean age calculation in order for methods to be sufficiently understandable and repeatable by readers.
- The visualization and description of the profiles in each figure could be improved upon, and separate AirCore profile data figures could be merged into one multi-panel figure for comparisons of species' profiles and their corresponding description within the text.

Specific Comments:

Figure 1: I am wondering where each package was located on the gondola. Could some packages in this figure be somehow labeled to indicate where measurements were made?

Table 1:
- Five – L82? (or three – L118?) different AirCores were used here, but it would be nice to have these (with perhaps IDs for each AirCore and dimensions, configuration, dryer, no dryer, etc.) described in Table 1 explicitly. Please also add ambient pressure and temperature sensors to Table 1 as these are equally important measurements.
- LISA and the FZJ AirCores are included in this table, but data are not presented. This is contrary to the approach taken in Table 3. Pico-STRAT Bi GAz is also not added in Table 1. I would suggest either adding all, but with "*" for measurement artifacts and/or descriptions in other, cited papers, or remove otherwise.

L125: Not all AirCores (e.g. FZJ) were analyzed by continuous-flow analyzers. Please correct/clarify. Were similar fill gas mixtures used for each AirCore?
L139/L159: It seems that the GUF and RUG AirCores samples were processed somewhat differently, from fill/push gas mixing corrections to the altitude attribution of the samples.

Please describe in more detail what the differences are between both GUF and RUG methods so that this can be assessed better.

Section 2.3: You might consider a 2.3.1 'GUF AirCores' vs. 2.3.2 'RUG AirCores' for organization

L169-L172: There is no mention of an ambient water vapor measurement onboard the gondola. It is difficult to examine water effects on the profiles (either dried or undried) without this type of measurement. For example, if the atmosphere is dry, no effects will be seen, so some clarification here and/or below in the description of differences seen between these profiles should mention something about the average tropospheric water vapor content above Kiruna.

Please define $Mg(ClO_4)_2$. Please also describe in more detail the altitude attribution used here. Is this similar to Tans, 2022 or Wagenhauser et al., 2021? How do these methods compare/differ?

Table 3:
- The text mentions that this table describes precision for both flask and AirCore measurements (L240), so the caption should mention this.
- It is a bit confusing with some species mentioned in this table. COS is mentioned for the RUG QCLS, but Table 3 precisions and results are not described. FZJ analysis precision by GC-MS is stated in Table 3 for only SF6 results with other species presented elsewhere. Please see comment above for Table 1.

Section 2.5: Could you please clarify the progression of air sample analysis and how many flasks were analyzed by each organization? Are all flask samples analyzed for halogenated species at GUF by GC-MS, by GC-ECD for CFC-12 and SF6, and by CRDS for CO2, CH4 and CO and then subsequently transferred to RUG and FZJ for additional species analysis?

Section 2.6: Please describe in a bit more detail how *AoA_from_convolution* works for the reader within the text. It might also be clarified in subsequent text that <u>mean</u> ages are calculated and used.

L271: Combining CO2, CH4, N2O profiles in a single figure would be advantageous to your description of the profiles within the text, and would allow the reader to better compare/contrast the variability between profiles.

Figures 2-6:
- It is very difficult to distinguish the colors between RUG AirCore 'wet' and 'dry'.
- Cryosample and other markers overlap each other at times and it's difficult to see what's underneath of these. Consider decreasing marker size here.
- Line width on PICO data covers AirCore profile data as well; consider reformatting so that all profiles can be sufficiently visualized.

L274: The altitude registration offset looks to be greater than 300 m (maybe 500-800m?), hence my request for additional description of the altitude attribution techniques above and potentially additional explanation for this offset in L281.

L303: The AirCore averaging kernel or diffusion is stated multiple times throughout the text. Could you include calculated uncertainties for this either in the text or in the figures above to indicate how much of an uncertainty the AirCore storage diffusion has?

L314: Can you offer any thoughts as to why AirCore CO_2 (and N_2O) seems to be lower than the cryosampler data of the same species? Could it be a problem with one or either of these sampling methods?

Technical Comments:
L17: 20 km 'MSL'?
L21: It is perhaps important to note that balloon data are also needed to validate and "calibrate" remote-sensing instrumentation like those in the TCCON and NDACC, not only to supplement them.
L24: What type of balloon platform (large or small-volume balloon) and instrument (flask, in situ) was used here? Please describe, as AirCore could potentially be used as a low-cost option that refutes the statement in L25.
L29: Please define chemical species names of CO_2 and CH_4, but also elsewhere in the manuscript for completeness.
L33: The AirCore approach, providing higher resolution profiles and additional air sample measurements, seems like an improved method over discrete flask sampling methods. Consider replacing the word "complementary"
L39: What does HEMERA-TWIN stand for, if anything?
L40: What are 'reference methods'? It might be more accurate to say that the gondola instrumentation allowed simultaneous comparison of several different measurement methods and not to say that one is more accurate than the other.
L42: Please describe why calculation of the mean age of air is significant and useful in a bit more detail for context.
L49: 'monotonic'
L59: ..bag sampling, [and] air sampling by means of …
L75: an average 'ascent' rate … and 31.2km 'MSL'?
L82: Please define BONBON acronym
L86: Please define LISA
L94: In Pico-STRAT Bi GAz, what does STRAT, Bi and GAz signify?
L133: Here and throughout the manuscript, text switches from spelling out 'Frankfurt', or 'University of Frankfurt' and 'Groningen' vs. using the acronyms for each organization. Consider using just acronyms for simplicity.
L149: part of the AirCore [CO_2 and CH_4?] profiles can be used…
L153: these are reported as dry mole fractions
L171: in subsection [space] 2.5 (Vincović et al. 2022; Tong et al, 2023).
L174: ..and [are] visible?

L185: upper and lower parts, or just upper parts?

L290: Above states that data is only collected upon balloon descent, so please correct.

L293: Can you please state over how long of a sample collection time this averaging takes place?

L308: The fine structure here is not seen in AirCores due to the diffusion of the samples, presumably.

L309: Delete "University of Frankfurt,", "by", and add on average, [by]

L326: age of air 'values'

L335: value derived 'from'

Figure 7: The x-axis label should be years, correct? Caption: "..markers represent 'Age' of air"

L372: Does the NOAA product come with a citation reference?