Schuck et al., The interhemispheric gradient of SF6 in the upper troposphere

Response to comments by Anonymous Referee #1

We thank the anonymous reviewer for the thorough reading of the manuscript and their suggested improvements. All comments are addressed in the following with the reviewer's comments printed in blue, and the responses in black.

General comment

I believe that this work has been achieved on the basis of the long history of experimental advances in balloon observations by the European scientific community. I suppose that earlier works such as Schmidt et al. and Fabian et al. have paved the way. It might be interesting if the authors provide readers opportunity to follow the origins of the active techniques and historical development. One additional paragraph in the introduction section could help. Technical advantages over the balloon observation systems from the US and Japan teams might be also interesting.

To acknowledge historic instrument development and balloon observations, the introductory sentence was extended with the following paragraph:

"High-altitude balloons continue to be the only means for in-situ observations of chemical composition at altitudes that cannot be reached by aircraft, i. e. above ca. 20km. Lightweight instrumentation, such as for example ozone sondes, can be lifted with small weather balloons, that can be launched routinely. Many trace gases, however, can only be measured with more complex instruments or from sampled air analysed post-flight in the laboratory (Ehhalt 1980, Fabian 1981). Cryogenic air sampling is an established method for the efficient collection of air samples in the stratosphere (Ehhalt 1974, Lueb et al. 1975, Schmidt et al. 1984, 1987) to obtain observations of trace gas profiles from the stratosphere."

Our measurement report does not provide advances in balloon observation systems per se but focuses on the comparison of established observation techniques for greenhouse gases that could also be deployed on other airborne platforms. Therefore, we refrain from comparing with balloon-based observational efforts in the US or Japan or other regions.

Age of air calculation should be elaborated so that readers can understand principles of the method better. Although I understand that this study stands on the past technical efforts, needed is more exact explanations about the method, not the software. The package should assume an age spectra formulation to find the best-match mean age of air. It could be more clearly explained why the surface reference time series of CH4 is used for calculation of the mean ages from CO2 and SF6 (were they or stratospheric data used for the CO2 correction?). What kind of assumptions brought the bias correction applied to SF6-derived ages of air might be also more explained with additional sentences, in addition to reference to Garny et al. paper. Although they are complex, the authors should provide essential parts. Since age of air is not uniquely determined from the data, I think the manuscript should provide the method in a way independently understandable.

The description of the age of air calculation was extended to make it understandable more easily without having to refer to the Garny et al. references. Please see response to specific comment on subsection 2.6 below for details.

Figures 2, 3, 4, 5 and 7 might be merged into a multi-panel graph. I am interested in comparing vertical profiles of different species particularly for correspondences of gradients among species. The tropopause could be marked, otherwise readers wonder at places like "in the troposphere." Profiles down to the troposphere are great advantage of this study.

The tropopause has been added to all profile figures as suggested (see below for further details). Merging all profile graphs into one figure creates fairly small panels. We have instead merged Figures 2 and 3, showing CO2 and CH4, and Figures 4 and 5, showing N2O and SF6, into one figure, respectively.

Specific comments

Throughout the manuscript: I've been wondering what HEMERA and TWIN stand for.

HEMERA is actually not an acronym, but the programme name refers to the ancient Greek goddess Hemera, the mythological personification of daylight.

The name TWIN of the gondola refers to the symmetric structure of the gondola frame. We have added this information to the text and to the caption of Figure 1. Although the name of the gondola may not be relevant for the discussion of the flight, but it helps to distinguish the payload from other HEMERA payloads launched in August 2021.

P2 L51: Though with some difference in methods, use of halocarbons for estimating mean age of air can be further back in time e.g., Daniel et al. (1996) and Harnisch et al. (1999). It might be worth mentioning that the idea existed much earlier than the recent extensive studies like Ray et al. (2017) and Leedham Elvidge et al. (2018).

In addition to the already cited work by Volk et al. (1997), we have added the suggested references in the revised version of the manuscript.

P3 L63: I could not find the companion paper (Laube et al. 2024) available. It must be made available before considering the decision of this manuscript. Otherwise this manuscript should provide more information without mentioning the reference.

Unfortunately, the processing of the companion manuscript had been significantly delayed. Laube et al. 2024 has thus changed to Laube et al. 2025, and it is now under review for the journal Atmospheric Measurement Techniques. In the meantime, the preprint has been published on EGUsphere:

> Laube, J. C., Schuck, T. J., Chen, H., Geldenhuys, M., van Heuven, S., Keber, T., Popa, M. E., Tuffnell, E., Vogel, B., Wagenhäuser, T., Zanchetta, A., and Engel, A.: Vertical distribution of halogenated trace gases in the summer Arctic stratosphere determined by two independent in situ methods, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-4034, 2025.

The reference in the manuscript will be updated correspondingly.

P4 L84: “position data” do they include the GPS altitude? Regarding this, the altitude coordination in this study is unclear. Are the altitudes in Figure 2 and the following figures the GPS altitude recorded by the Pico-SDLA instrument? Or was pressure recorded and converted to altitude?

The position data include GPS altitude, latitude and longitude from the GNSS system onboard Pico-STRAT Bi Gaz. Pressure measurements were provided by Pico-STRAT Bi Gaz, that uses a ParoScientific Inc absolute gauge to measure air pressure at high accuracy (0.01%).

The information in the payload subsection is extended correspondingly in the revised version and now reads:

“Position data as well as ambient pressure and temperature were recorded by the Pico-STRAT Bi Gaz instrument. The position data include GPS altitude, latitude and longitude from a GNSS system, pressure measurements are performed with a ParoScientific Inc. absolute gauge, and ambient temperature values are recorded by three fast-response temperature sensors (Sippican).”

Axis labels were modified to “GPS altitude [km]”.

P4 Section 2.2: It is regrettable that the Pico-STRAT Bi Gaz instrument for CO2 and H2O did not work well. Looking at Figure 2, I hope successful measurement at next opportunity as the present profiles were all from “reconstruction” from subsequent analyses.

The Pico-STRAT Bi Gaz instrument, which measures CO_2 and H_2O mole fractions was affected by an external RF modulation, which distorted the CO_2 spectra. Water vapour measurements seem much less affected and the measurements are still valuable, but were not used.

Nevertheless, since the instrument was flying inside a larger gondola, water vapour measurements, are contaminated by the outgassing balloon wake from the cold point tropopause. The amplitude of this contamination slightly increases from 5 to 15 ppmv as the balloon ascent in the stratosphere, due to the increased number of outgassed water vapour molecules. The natural mean stratospheric water vapour mixing ratio is around 5 ppmv at high latitudes. In the troposphere, the natural mixing ratio dramatically increases by 4 to 5 orders of magnitude, which renders the contamination negligible.

This issue is well known and in dedicated campaigns, focusing on stratospheric water vapour studies, the hygrometers are located far away from the balloon (distances higher than 30 m) to reduce the contamination effect in the stratosphere. In the frame of this flight, measuring inside the gondola was done since the instruments were flown in the contaminated environment. Please note that the CO_2 profile from Fig.2 comes from AirCore measurements, while CH_4 profiles from Pico-STRAT are original profiles without any reconstruction.

P5 Table 1: I think that this table could be somehow merged with Table 3, because I have sometimes confused with which instrument for which species. A merged table might provide a quick look for such correspondences.

Merging Table 1 into Table 3 has turned out to result in a difficult to read table. We therefore repeated some information from Table 1 in Table 3 to make it easier to relate the given information to the text. Please note that following a suggestion by the second reviewer, an additional table, comparing technical details of the different AirCores, was included in the revised version of the manuscript.

P8 L192: Is there a reference that described usefulness of the cotton filter for COS measurements? Does co-existence of ozone and COS deteriorate their stability in the stainless steel flask? Such an explanation is valuable and desirably presented in this manuscript or the companion one.

Andreae et al. (1985), Hoffman et al. (1992), Persson and Leck (1994), Engel and Schmidt (1994) described the risk of possible biases due to reactions of ozone with reduced sulfur compounds and/or other tracers, both at tropospheric and stratospheric altitudes. Hoffman et al. (1992) and Persson and Leck (1994), in particular, focused on cotton scrubbers for tropospheric sampling. Engel and Schmidt (1994) observed depleted COS without ozone removal. More insights will be presented in Zanchetta et al. (in preparation, 2025).

All publications mentioned here are included in the references of the manuscript.

P8 Section 2.5: I think that it would be valuable to describe how much sample amount (volume) was used for each analysis. This information should be critical when the campaign was planned.

The Frankfurt GC-MS uses 1l per analysis, each sample was analysed twice. The QCLS runs for 5 minutes at 50sccm amounting to 250 ml. The CRDS uses 35ml/min over 180 s, adding up to 105ml, the GC-ECD setup used approximately 270 ml. The GC-MS analysis at FZJ used 250ml.

This information has been added to the manuscript upon revision extending the first paragraph of subsection 2.5. Air Sample Analysis:

"All flasks of the cryo sampler were analysed at GUF using gas chromatography/mass spectrometry (GC-MS) for halocarbons, by gas chromatography/electron capture detection (GC-ECD) for CFC-12 and SF6, and with the continuous-flow CRDS used for AirCore analysis for CO2, CH4 and CO. The sample volume used for these analyses was two times 1 L for GC-MS, and approximately 0.27 L and 0.11 L for GC-ECD and CRDS analysis, respectively. The cryo sampler was then transferred to RUG where the continuous-flow QCLS used for AirCore analysis was deployed to measure CO2, CH4 and N2O from the sample flasks. This used approximately 0.25 L of sample volume. Last, samples were analysed for SF6 at FZJ using a GC-MS setup that needs 0.25 L of sample volume."

P9 L219: How long does it take to have the instrument output stabilized with the flow rate ~50 sccm. I imagine that those stable data points were taken to calculate the measurement value and the time for stabilization determined the amount of sample used.

The following paragraph was added to the next to discuss the aspect of instrument stabilization for the QCLS:

"The instrument stabilization is of 'double exponential' character, i. e., exhibiting an initially rapid approach to the final value, but then taking a long time to become stable. This depends on species and is more pronounced for CO2 than for CH4. The measurement duration of ~5 minutes, chosen to conserve precious sample, does not for all samples result in quantitative replacement of the previous sample. In order to maximize the value of our analysis, a double exponential function was fit to the measurement data and used to predict the true sample value, i.e., the asymptote of the function. The deviations between the asymptote and the mean of the last 60 seconds of a sample was for almost all samples smaller than the uncertainty of either. That means that this method does not make unjustified assumptions and does not add significant uncertainty to the results."

P10 Section 2.6: As in my earlier comment, I hope this section be reformulated so that readers can correctly follow principles, assumptions and processes to calculate the age of air.

The section has been extended for the revised manuscript as follows:

"2.6 Age of air calculations

The mean transit time that it took for all contributions to an observed air parcel to arrive at the observation location from their respective entry points into the stratosphere is defined as the mean age of air. Thus, the observed mixing ratio of an inert trace gas at some place in the stratosphere is determined through the distribution of transit times, called the age spectrum, and the long-term change in its mixing ratio at the entry point. Commonly, the age spectrum is mathematically described as an inverse Gaussian function, and the mean age of air is the first moment of this distribution (Hall and Plumb, 1994; Waugh and Hall, 2002). Mean age of air values were derived from SF_6 measurements and independently from simultaneous CO_2 and CH_4 measurements following the procedure described by Garny et al. (2024b) using the AoA_from_convolution *python package version 1.0.0 (Wagenhäuser et al., 2024). In brief, this method calculates the expected mixing ratios of an inert trace gas through a mathematical convolution of the age spectra for different mean age values and the mixing ratios time series at the entry point. The mean age is then determined from the best match between the observations and the mixing ratios resulting from this forward calculation. The derived mean age is the age value for which the convolution best fits the observed mixing ratio.*

To calculate mean age from tracer observations, also the time series of the respective tracer at a reference surface is needed. For this purpose, the AoA_from_convolution *package uses the NOAA Greenhouse Gas Marine Boundary Layer Reference for SF_6, CO_2 and CH_4 trace gas mixing ratio time series at the tropical surface ±17.5° around the equator (Lan et al., 2021; Garny et al., 2024c). The inverse Gaussian describing the age spectrum is parameterised assuming a ratio of moments as described in Garny et al. (2024b). Mean age values below 1 year are omitted due to numerical reasons of the software implementation. Regarding SF6, these mean age alculations do not account for the mesospheric sink, which leads to apparently older SF_6 mean ages (Leedham Elvidge et al., 2018; Garny et al., 2024a).*

For CO_2, the software first uses CH_4 mixing ratios to account for stratospheric CO_2 production from CH_4 degradation. This corrected CO_2 mixing ratio is then used to derive the mean age. Note that the seasonal cycle of CO_2 propagates into the lower stratosphere, and it is impossible to disentangle the seasonal cycle from the long-term increase. Therefore, mean age values below 2 years derived from CO_2 measurements are problematic (Garny et al., 2024a)."

P10 L265: "in the troposphere" the tropopause should be presented. Otherwise readers cannot be clear about which altitude range is mentioned.

We agree that adding the tropopause helps to read the graphs and have added it to all profile plots. We have derived the tropopause height from the radiosonde data during the balloon ascent by calculating the slope of the temperature profile. The tropopause height has been assigned at 10.5km where the temperature change with altitude first reached the value of 2K/km following the WMO definition. A corresponding explanation has been added to the revised manuscript.

P14 L303: It is interesting to see the ups and downs in CH4 above 20 km observed by the Pico instrument. Is there any possible mechanism that could shape such layers?

The origin of such structures is under investigation, however, we do not have an explanatory hypothesis and believe that this is beyond the scope of a measurement report. Some of the features are associated with ozone structures, observed by an ozone sonde flown 2 hours later than the ZPB flight of TWIN which is analyzed here. Similar structures were also observed during an earlier flight of the Pico-SDLA instrument from Kiruna in 2016.

These structures were also resolved with the sub-sampled MegaAirCore and this is discussed in more detail in the companion paper by Laube et al. 2025.

Technical Comments

P4 L70: “known” to “considered”

Changed as suggested.

P11 L274: “caused”

P18 L350: “Aircore” to “AirCore”

P18 L356: “altitde" to “altitude”

P18 L358: “Age of Air” to “age of air”

These typos have been corrected in the revised version of the manuscript.