

Reply to Interactive comment of Referee #2 on “LISA: a lightweight stratospheric air sampler” by Joram J. D. Hooghiem et al.

We thank the two reviewers for their helpful and detailed comments on our manuscript. We have addressed the major comments from both reviewers regarding our derivation of the bias estimation in isotopic composition measurements and revised the introduction according to the comments of both reviewers. Below, a detailed point-to-point reply and a revised version of our manuscript with track changes are given.

General comments

The authors present a new stratospheric sampling system that can fill a niche between current techniques, as it provides larger air samples with better vertical resolution than AirCore while operating with much lower cost and payload weight than typical cryogenic whole air systems. The manuscript is suitable for AMT and well-written and should be published after addressing the following comments and requests. In particular, a possible error in the calculation of uncertainty for isotopic analysis should be resolved.

Major issue: Section 6.4

I agree with Reviewer 1 that the underlying assumptions used to estimate fraction of sample from contamination (f_c) are incorrect. The ultimate source of the error seems to derive from the expression used to define $f_c = \text{bias} / ([X] + \text{bias})$. Here I presume that $[X]$, defined by the authors as “typical mole fraction,” is therefore meant to be $[X]_{\tilde{m}}$, the measured mixing ratio.

If we rearrange equation (7) to $\delta m = \delta s - \Delta\delta$ and rewrite for mixing ratio as $[X]_m = [X]_s - \Delta X$, where $[X]_s$ is the stratospheric mixing ratio and ΔX is the bias, then substitute this into the equation above, we have $f_c = \Delta X / (\Delta X - [X]_s + \Delta X) = \Delta X / [X]_s$. But since the authors also define $f_c = \Delta X / ([X]_s - [X]_c)$ in equation (8), this implies that the authors are assuming the mixing ratio of the contamination is 0 when they calculate the f_c on pg 14, line 28. Instead, they should use measured or typical mixing ratios, nominal 400 ppm and 1800 ppb for CO₂ and CH₄, respectively. This would result in roughly the same estimate as Reviewer 1 has calculated.

We would like to refer to the reply the comments of reviewer 1. We have copied it here for convenience.

We thank the reviewer for pointing out the calculation error. We mistakenly assumed that the contamination results from one-way diffusion of ambient CO₂ into the bag. If the observed bias would be entirely caused by diffused contamination air, the fraction of the contamination air is indeed about 8 percent. However, the observed bias is unlikely caused by diffusion alone based on our laboratory experiment results, where we found much smaller biases (~0.1 ppm for CO₂ and ~2 ppb for CH₄), as shown in Figure 2.

We have further thoughts on the diffusion process. According to the Fick’s law, the diffusion depends on the diffusivity and the concentration gradient, which may be different for different species, e.g. the concentration gradient for O₂ and N₂ across the bag’s material is negligible, and the concentration gradients for CO₂ and CH₄ depend on their ambient concentrations. We have updated the manuscript with the following analysis of the problem.

“The stratospheric air samples can be used for analysis of isotopic composition measurements of trace gases. Here we take CO₂ and CH₄ as an example to estimate the uncertainties of isotopic composition measurements due to the storage bias (see Table 3) or the AirCore-LISA bias (see Table 5), and the estimated isotopic signatures associated with the assumed contamination source.

For any species, the measured number of molecules n_m , is the sum of the number of molecules from the original source, n_s , and the contamination n_c that entered the sampling bag through diffusion:

$$n_m = n_s + n_c \quad (6)$$

Diffusion is governed by Fick’s law:

$$J = D \frac{\partial C}{\partial x} \quad (7)$$

Where J is the diffusion flux, D the diffusivity and $\frac{\partial C}{\partial x}$ the concentration gradient of diffusing species, e.g. CO₂. Each layer of the MLF bag has its own specific diffusivity and sorption characteristics, and for simplicity here we assume that it can be modelled with a single diffusivity constant that is only species-dependent. The process of sorption of gases into the solid material and the interaction governing that process is also ignored for simplicity. Then the total diffusion out of the bag which is of stratospheric origin, denoted by y_{out} . The total amount diffusing into the bag, denoted by y_{in} , is of tropospheric origin. Assuming that these two fluxes can be modelled with a law of mass action:

$$y_{in} = DC_c \quad (8)$$

and

$$y_{out} = DC_s \quad (9)$$

Where C_c is the concentration of ambient air that contaminates the stratospheric sample in the bag denoted by C_s . The bias that we measure is as follows:

$$bias = y_{in} - y_{out} \quad (10)$$

Assuming that y_{out} is purely stratospheric and y_{in} is purely tropospheric, and that c_{out} and c_{in} are mole fractions outside and inside the bags and are assumed to be constant. We can estimate the fraction of tropospheric contamination in our sample. The measurement (C_m) is the sum of the original stratospheric sample (C_s), minus the sample that is leaving the bag and the tropospheric sample entering the bag:

$$C_m = C_s - y_{out} + y_{in} \quad (11)$$

And the contaminating fraction is then f_c is

$$f_c = \frac{y_{in}}{C_m} = \frac{C_c \cdot bias}{C_m(C_c - C_s)} \quad (12)$$

Where the right-hand side is obtained by using Eq. (8) to (10). The sample fraction, f_s , is simply

$$f_s = 1 - f_c \quad (13)$$

The isotope composition after the mixing of the tropospheric contamination into the sample air, can be approximated with:

$$\delta_m \approx \delta_s f_s + \delta_c f_c \quad (14)$$

were δ_m is the final isotopic composition, and δ_s and δ_c represent the isotope composition of source and contamination and f_s and f_c are the fractional contributions to the total number of molecules after mixing. We further define the bias of the isotopic composition measurement as

$$\Delta\delta = \delta_s - \delta_m \quad (15)$$

Combining Eq. (8) and Eq. (9), we derive

$$\Delta\delta = (\delta_s - \delta_c) f_c, \quad (16)$$

For the calculation, we regard the mean differences between AirCore and LISA measurements (Table 5, e.g. 0.84 ppm for CO₂ and 1.8 ppb for CH₄) as the upper limit of bias induced in the stratospheric samples. Another estimate is performed based on the storage test results, that showed maximum drift of 0.11 ppm CO₂ and 2 ppb CH₄, presented in Table 3.

The fraction f_c can be calculated according to Eq. (12) with C_s being the typical stratospheric mole fraction, which is taken to 395 ppm for CO₂ and 500 ppb for CH₄. We use typical tropospheric values of 405 ppm for CO₂ and 1800 ppb for methane. The isotopic compositions δ_c and δ_s are taken from various references, and are presented in Table 7. The resulting bias in the isotopic composition measurements are presented in Table 7.

We can readily see that the estimated uncertainties due to the storage bias are relatively small compared to the typical analytical precisions, also presented in Table 7. Hence the LISA sampler provides a viable sampling tool for accurate measurements of stable isotopes in CO₂ and CH₄.”

Table 7: Expected bias in stable isotope measurements on samples obtained by LISA, due to the limited accuracy of the LISA sampler. Typical values for the troposphere and stratosphere are taken from the indicated references: A) (Troler et al., 1996) B) (Mrozek et al., 2016) C) (Nisbet et al., 2016) D) (Bergamaschi et al., 2001) E) (Aoki et al., 2003) and F) (Röckmann et al., 2011). Reported measurement reproducibility's, R_e , for stratospheric air are also provided. $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values are with respect to Vienna Pee Dee Belemnite (VPDB) and $\Delta^{17}\text{O}$ and $\delta^2\text{H}$ are with respect to Vienna Standard Mean Ocean Water (VSMOW). f_c was calculated using a source value 395 ppm (CO₂) and 500 ppb (CH₄). For f_{c1} contamination values of 0.84 ppm (CO₂) and 1.8 ppb (CH₄) based on LISA AirCore observed mean bias, resulting in $\Delta\delta_1$; For f_{c2} the maximum observed drift (Figure 2) of 0.11 ppm (CO₂) and 2 ppb (CH₄) are used, resulting in $\Delta\delta_2$.

Species	$\delta_c(\text{‰})$	$\delta_s(\text{‰})$	$R_e(\text{‰})$	LISA-AirCore		Storage Test	
				f_{c1}	$ \Delta\delta_1 (\text{‰})$	f_{c2}	$ \Delta\delta_2 (\text{‰})$
$\delta^{13}\text{C}$ (CO ₂) (VPDB)	-7.5 (A)	-8.4 (E)	0.02 (E)		0.077		0.01
$\delta^{18}\text{O}$ (CO ₂) (VPDB)	-2 (A)	12 (E)	0.05 (E)	0.086	1.203	0.011	0.158
$\Delta^{17}\text{O}$ (CO ₂) (VSMOW)	0 (B)	7 (B)	0.2 (B)		0.602		0.079
$\delta^{13}\text{C}$ (CH ₄) (VPDB)	-47 (C)	-20 (F)	0.7 (F)	0.006	0.134	0.006	0.149
$\delta^2\text{H}$ (CH ₄) (VSMOW)	-85 (D)	190 (F)	2.3 (F)		1.366		1.517

An alternative method to estimate f_c would be to solve equation (8) for both CO₂ and CH₄ simultaneously. Since the authors already have estimates for $\Delta\delta$ and δ_s stated (pg 14, lines 26-28), this results in two equations with two unknowns, which can then be solved trivially. This would be a worthwhile exercise to confirm the estimate of f_c provided by Reviewer 1.

Since the diffusion through the bag's material is species dependent (e.g. species dependent diffusivity and concentration gradient.), the fraction f_c is also different for each species. That is why we find different values for f_c as presented in table 7. In a simple mixing scheme with two air masses the suggested calculation would indeed be true.

Finally, it is not clear to me why the authors don't use the storage test results, in addition to the ambient data, to estimate f_c . This should either be performed, or the authors should explain why this analysis is not valid.

In the revised manuscript we have added the estimate based on the storage test results.

Additional general comments

I agree with Reviewer 1 that the introduction is a bit unfocused and broad. I think a more focused discussion of the literature with regards to the particular challenges of making sufficiently accurate and precise mixing ratio measurements in the stratosphere would be more useful. See the introduction offered by Membrive et al., 2017 for an example.

The manuscript presents multiple tables of information (e.g. Table 2 is derived fit coefficients, Table 3 is instrument operational settings) that would be better suited in supplemental materials, as this information is not critical to main discussion here). Meanwhile, there is no table provided that summarizes the instrument specifications (e.g. weight, power requirements, sample resolution at stated altitude) and comparison to the AirCore and other systems. I would be appreciative of such a summary table.

We have added a table summarizing the key components of the sampler in section 2. We like to draw the attention to table 5 where a comparison of sample size and resolution between AirCore and Sampler, which to our best knowledge are to only to instruments with weight lower than 5 kg that can sample from the stratosphere.

In the light of the discussion presented in sections 6.2 about sampling size and vertical resolution we deemed the information in Table 3 relevant. Table 2 is removed from the manuscript, as we agree I does not provide any relevant information.

“Table 1: Components used in the LISA sampler, including manufacturer and product key. The total weight is given for amounts per part. Voltage and power are presented according to manufacturer specification. The total weight for the onboard computer and sensors is given.”

Component	Company	Product key	Amount	Voltage (V)	Power (W)	Weight (g)
Servo motor	Hitec	HS-65HB+	5	4.8-6	1.32	91
Pump	KNF	NMP 850.1.2 KNDC B	1	24	10.8	403.6
Bag (MLF)	Supelco	30227-U	4	(-)	(-)	80.4
Tube	Cole Palmer	EW-95100-02	1	(-)	(-)	30
Union T	Swagelok	NY-400-3	5	(-)	(-)	39
Union Knee	Swagelok	NY-400-9	5	(-)	(-)	33
Battery	(-)	CR123	10	3	(-)	166
Pressure sensor	Honeywell	HSCMAND015PASA 5	2			
Temperature sensor	IST	600C (100Ohm)	1	7-12	(-)	87.4
Datalogger	Arduino	Mega 2560	1			
Battery Casing	TruPower	BH-CR123A	10	(-)	(-)	68.8

Additional comments

p. 1, line 11 - The abstract discusses the stability tests, which include H₂O mole fractions. However, the rest of the abstract only discuss CO₂, CH₄ and CO mole fractions. A phrase noting the significance of H₂O (interferent?) would be appreciated.

We have removed H₂O from the abstract. H₂O was measured and used to obtain dry mole fractions of CO₂, CH₄ and CO, as mentioned in Section 3.1. We are interested in the stability of H₂O in the bags as well, due to its potential effect on isotopic compositions of CO₂. H₂O mole fraction measurements are now only left as a discussion in section 3.

“H₂O measurements are used to obtain dry mole fraction of CO₂, CH₄ and CO using the water vapour corrections described in Chen et al., 2013 and Rella et al., 2013, before assessing drift of these species.”

p. 1, line 29 – “Stratospheric changes in ozone and water vapour levels” is awkwardly phrased. How about “Changes in stratospheric ozone and water vapour levels”?

We have adopted the suggested phrase.

p. 2, line 1 – “stratospheric air up to 35 km” Please define this value (e.g. ASL).

“Above mean sea level (a.m.s.l.)” has been added.

p. 2, line 13 – “remarkable scientific efforts” I would avoid using words like “remarkable” without a substantial defence of this term.

We have adjusted the sentence and avoided the term “remarkable” care has been taken to avoid such terms throughout the text.

p. 2, line 15 – Introducing mean age of air is unnecessary here unless it is discussed later in the text.

We have left out the discussion of mean age since it is indeed not relevant to this manuscript.

p. 3, line 8 – “AirCore does not provide large sample amount” Please quantify this statement.

We have added typical values of 300 to 600 mL of stratospheric air (200 to 0 hPa) sampled with AirCore.

“The volume of air sampled between 0 to 200 hPa (12 to 30 km) by the AirCore ranges from 300 to 600 mL, depending on the geometry of the AirCore.”

p. 4, line 1 – “a diaphragm pump (KNF, product no. NMP 850.1.2 KNDC B)” Please provide a description of wetted surfaces for this pump.

A description has been added in the manuscript, after personal communication with KNF.

“The pump utilizes an EPDM rubber diaphragm (35 mm diameter) and valves, and a small piece of flexible PU tube.”

p. 5, line 20 – Storage test results. Can the authors take advantage of the results of these tests to separate leak and permeability effects for all species?

Diffusion through the bags material is species dependent. Furthermore, the diffusivity through the materials is unknown, and determination of the diffusivities require experiments beyond the scope of this work. Therefore, the experiments do not allow to separate leak and permeability effects. Besides these, most of the time the pressure difference between in and outside the bag is 0, hence there is no driving force for air to leak in or out. Only during ascent, the pressure inside the bag is higher than ambient and air might leak out. Thus, diffusion

through the material and possibly through the septum of the inlet is likely the main cause of the observed drift in mole fractions during the storage tests.

The discussion of the water vapor experiment is nicely done and touches on observed changes in CO; can this be extended?

With our current knowledge and data, we can only speculate the different reasons for the observed changes in CO. In the manuscript the only remaining correlation we could find from the data is that with water vapour. At this stage we can only postulate a potential bias in the water vapour correction function. Though the water vapour corrections are well quantified, they are not tested with low mole fractions.

p. 6, line 11 – “we observed an offset” Please describe which direction these offsets were (increase or decrease). As a reader, I wondered if this effect could have been a cause of the CO₂ offset observed during the April flight discussed later.

In the revised manuscript the direction, which is positive, has been clarified. There is no difference in preparation between the April and September flights. Due to unavailability of N₂ the bags were preconditioned with our calibration gas in the field prior to each flight. If the offsets would be the cause, it should have been most notable in CO, with values higher than AirCore, which is not the case. We therefor conclude that the offsets prior to preconditioning contribute little to the observed AirCore-LISA discrepancies.

“We found out that it is necessary to precondition the MLF bags before use, because we observed a positive offset of ~12 ppm CO₂, ~8 ppb CH₄ and ~30 ppb CO between the immediately analysed results of un-preconditioned MLF bag samples after filling and the assigned cylinder values”

p. 6, line 24 – “We do not include any sampling error in the presented evaluation, although it might be significant for high-precision measurements, especially for CO in the stratosphere when the ozone concentration is high.” This statement was ambiguous to me. Please explain exactly what is meant by “sampling error” and the significance of ozone concentration.

At the time of writing, we hypothesized that ozone might be reacting with the pump materials. However, this was checked later and apparently ozone concentrations of up to 1000 ppm does not have a significant effect on the EPDM material:

http://www.ozoneapplications.com/info/ozone_compatible_materials.htm. The sentence has thus been removed.

The sampling error encompasses any contamination effects or reactions that might alter the mole fraction during the sampling process. This has been adopted in the revised manuscript.

“The sampling error encompasses any contamination introduced by the sampling system itself. This includes chemical production of the species of interest and residual air in any dead volumes of the manifold. The chemical production during sampling is likely to be very small for two reasons. First of all, the wetted surfaces, Kynar and EPDM diaphragm, are chemically inert. Secondly, the high flow rate minimizes exposure of the sample to materials used in the sampler and hence chemical interaction with the wetted surfaces is limited. In addition, the flushing procedure with high flowrates ensures multiple turnovers of the manifold, which

reduces the surface effects on the sample. These effects are thus assumed to have no influence on the CO₂, CH₄ and CO mole fractions.

The dead volumes in the manifold are a potential source of contamination bias. The dead volume is estimated to be 1.5 mL per sample and will be at local ambient pressure prior to sampling. So, the dead volume uncertainty, σ_v , to the contribution to a sample at 200 L_{stp} sampling is very small."

p. 6, line 28 – "However, detailed information about storage conditions are required to correct for the drift, which is usually unavailable in the field." Why wouldn't this be available? The detector used for analysis was in the field (see p. 9, line 27) so this data may have been collected for the samples presented here. Please explain if the data is available and whether it should be collected in future work.

The statement in P. 9, Line 27 (original manuscript) was ambiguous. The samples were analysed in the FMI laboratory at the Sodankylä observatory, which to us, being from Groningen, is "in the field". The statement has now been modified in the revised manuscript.

"After retrieval of the payload, the samples were analysed in the TCCON laboratory"

The detailed storage information includes 1) ambient mole fractions of different species; 2) ambient temperature; 3) ambient pressure. The ambient temperature and pressure were logged by the datalogger, and the ambient mole fraction measurements of different species were not done as it would require a dedicated precise and accurate analyser. Since a robust correlation between the ambient information and the drift during storage has not been established, we would not recommend collecting this information in future work.

p. 7, line 12 – "The performance of the small diaphragm pump is to our best knowledge not previously investigated under the atmospheric conditions in the stratosphere, e.g. at low-temperature and low-pressure conditions." This is a nice presentation of this experiment. Was the pump subject to stratospheric temperatures? The manufacturers specification is for ambient temperature between +5°C and +40°C. Any insight that can be provided about the pump performance at stratospheric temperatures would be appreciated.

Indeed, no laboratory tests were performed at stratospheric temperatures. The pump performance could be affected by the cold environment. First, the batteries could lose capacity and cause the power supplied to the pump decreasing. The temperature inside the thermally insulated package, where the pump is located, during flight ranged between 30 and -15 °C. Secondly, the diaphragm is exposed to the cold air passing through the pump. The elasticity of a rubber is temperature dependent, which could reduce the performance of the pump. On the other hand, heat is released during operation of the pump, which increases the temperature. Finally, the effect of air temperature on sample size follows the ideal gas law, and the sample size increases at low temperatures. As no experimental data is available to determine the performance of the pump at stratospheric temperature, we assume that the pump performs the same during flight as at room temperature for the calculation of the sample size. The best insight we have is the comparison of the flight results in Fig. 4.

We have clarified in the manuscript with the following text in section 6.

“During the experiments described in Section 4, the pump was at room temperature. The pump performance could be affected by the cold environment. First, the batteries could lose capacity and cause the power supplied to the pump to decrease. The temperature inside the thermally insulated package, where the pump is located, during flight ranged between 30 and -15 °C. Secondly, the diaphragm is exposed to the cold air passing through the pump. The elasticity of a rubber is temperature dependent, which could reduce the performance of the pump. On the other hand, heat is released during operation of the pump, which increases the temperature. Finally, the effect of air temperature on sample size follows the ideal gas law, and the sample size increases at low temperatures. As no experimental data is available to determine the performance of the pump at stratospheric temperature, we assume that the pump performs the same during flight as at room temperature for the calculation of the sample size.”

p. 8, line 6 – “19.7 seconds.” How precise is this value? I’m a bit surprised that the pumping speed would be so consistent at varying pressure. There are no data points shown in Figure 3 prior to 20 sec, which makes this statement a bit hard to evaluate. The uncertainty of this value under operating conditions should be discussed with respect to the modelled behaviour that follows.

Initially pressure is constant as the bag is expanding. When expanded to its full size, the pressure starts increasing, which was found to be around 19.7 seconds which gave the best fit. This was based on visual inspection of the data. We have added an estimate of the accuracy of this number of about plus or minus 1 second, since we cannot justify otherwise.

“The sampled air volume increases linearly with the sampling time when the bag expands to its full size during the first 20±1 second’s. The moment compression is required, pressure starts increasing rapidly, and this moment was found to be 20 seconds after filling initiated.”

p. 8, line 25 – “The sampled volume at STP can be modelled for all pressure levels ranging from 200 to 0 hPa” It seems inappropriate to model to 0 hPa, given that the data collected ends at 30 hPa.

We have rephrased “modelled” to “approximated”. This was ultimately the goal of the experiment. It provides us with a tool to determine the sampling strategy in the field. And the approximation suffices.

“Combining Eq. (1) and Eq. (2), the sampled volume at STP can be approximated for all pressure levels ranging from 200 to 0 hPa for any chosen sampling time.”

p. 9, line 5 – “To be on the safe side,” should be rephrased “Conservatively,”

The sentence has been rephrased according to the suggestion made by the reviewer.

“Conservatively, we set the maximum absolute pressure in the MLF bag during flight not higher than 280 hPa to avoid any potential loss of sample due to the burst of the bag.”

p. 11, line 10 – “the vertical resolution increases” The vertical resolution decreases since the vertical height increases. There are several significant deviations from the relationship between altitude and vertical resolution, as shown in Figure 4b: 26-Apr at 150 hPa, 4-Sep at 100 hPa and 5-Sep at 30 hPa. Could these data points be explained in the context of the discussion of the variability in vertical resolution starting at p. 11, line 14?

Indeed, the vertical resolution decreases with altitude. As explained in P. 11, Line 28 there were 3 factors that explained all the existing variability 1) varying sampling time; 2) varying ascending speed; 3) varying maximum allowed pressure.

We have highlighted the samples mentioned by the reviewer above in the manuscript.

“The variability in the vertical resolution is the result of three factors: 1) varying sampling time; 2) varying ascending speed; 3) varying maximum allowed pressure. The ascending speed was typically around $7\text{--}9\text{ m s}^{-1}$ in the lower stratosphere and decreased to $4\text{--}5\text{ m s}^{-1}$ in the middle stratosphere. The varying ascending speed accounts for the observed deviations from the otherwise linear trend in Fig. 4b. In the lower stratosphere (10 to 15) km the maximum allowed pressure inside the bags was usually reached in a period shorter than the pre-set sampling time, leading to relatively high vertical resolution. In the region 10 to 15 km two samples deviate (4-Sep 17 km and 26-April 14 km), with lower resolution, which is due to a higher ascending speed. In the middle stratosphere, the sampling time was usually the limiting factor to vertical resolution. One sample in the middle stratosphere has a relatively good vertical resolution (5-Sep, 25 km), which is due to the relatively slow ascent speed.”

p. 11, line 25 – While the vertical profiles of the mixing ratios are shown in Figure 5, I don’t see the actual values presented. Could this be added to Table 3 or as a Supplemental Table (along with individual uncertainties of both mixing ratio and vertical height)?

The data is available on request.

p. 12, line 8 - While there are significant differences between the LISA and AirCore retrieved mixing ratios for CO₂, the CH₄ and CO mixing ratios show good agreement. Can this allow the authors to make a statement about the validity of their assumptions for estimating altitudes and pressures for the failed data loggers? And does this indicate that the offset CO₂ measurement is a result of the measurement of mixing ratio rather than altitude? This comment is also for p. 12 line 30 to p. 13, line 2.

This is a very good suggestion, which we have added in our discussion (section 6.1). The CO is not a good diagnostic since virtually no gradient is observed in CO (within measurement uncertainty). CH₄ however is a good diagnostic and indeed the observed LISA and AirCore CH₄ profiles show that the altitude registration is reliable. The CO₂ offset must be related to a difference in measurement and cannot be attributed fully to altitude registration.

“The AirCore valve did not close during the 26 April 2017 flight. This complicates the altitude registration of the AirCore. However, the large degree of agreement in CH₄ profiles between AirCore and LISA shows that AirCore altitude registration is reliable for the flight on 26 April 2017. The bias in CO₂ cannot be attributed to a failure in altitude registration and hence must be related to mole fraction measurement.”

p. 12, line 17 – “0.5 ppm or 0.13%” I’m not sure that the percentage reported here is appropriate, as it implies greater analytical precision than is achieved. If the CO₂ mixing ratio varies only 20 ppm or less between the troposphere and the top of the sampling region, then 0.5 ppm is a more significant deviation. I would suggest simply removing these percentages.

We agree that depiction of a percentage leaves room for interpretation. The percentages are thus removed from the manuscript.

p. 13, line 13 – “decreases fast.” Should be “decreases quickly.” In the following sentence, “the gain in sample amount, for example adding 10 extra seconds of sampling time, does not increase” is not correct, but rather the gain in sample amount is very small in comparison to the decrease in vertical resolution. This should be restated.

The sentence in the original manuscript was indeed ill phrased and has been updated according to the reviewer’s comment.

“From the results shown in Fig. 3d, we see that after 200 seconds of sampling the gain in sample amount decreases quickly. Therefore, the gain in sample amount, for example adding 10 extra seconds of sampling time, is small, however the decrease in vertical resolution is significant.”

p. 13, line 16 - As an aside, I would suggest additional bags as an alternative to larger bags, as the altitude profiles presented in Figure 5 indicate that there are periods during the ascent when no sample is collected, i.e. the system is sitting idle.

We agree that additional bags would be a good option to increase the vertical density of the sampling, and the system presented in this manuscript can indeed be easily adapted for that. However, for other scientific needs, e.g. for more precise $^{14}\text{CO}_2$ measurements, larger samples are preferred. We have modified the text to include both cases.

“Alternatively, to increase the amount of sample retrieved during one flight, additional bags can be considered. Currently the system is idle during several stages of the ascent as can be inferred from Figure 5. This will however be more demanding on battery power. Furthermore, care has to be taken to avoid overlapping sampling schemes i.e. sampling of a sample at altitude P_1 is still ongoing while the set-point altitude for sample two, P_2 is reached. This is complicated further with variable ascent speed, that is typical for these balloon flights.”

p. 13, line 28 – It is not clear why the uncertainty in the sample amount is important. Could the authors explain the importance of this?

The sample amount is important because it often significantly influences the precision of the measurements by instruments. Since we aim to obtain larger sample amount and the sample amount was only determined indirectly from pressure and temperature measurements, we think it logical to estimate the uncertainty of the sample size. We have added the uncertainty to the captions of Figures 3 - 4.

p. 23 – Figure 3b. I found the reuse of marker shapes between 3a and 3b confusing (I was looking for a relationship between the same marker shape in both figures). Could the authors change the marker shapes in Figure 3b to avoid this? For Fig 3c, it is not clear if the units on the vertical axis use volumetric liters or standard liters. This should be standard liters for consistency. Also, in Fig 3d please extend the vertical line at 0.76 L for the full vertical height of the figure, as was done in Fig 3b.

We have adjusted the marker shapes so that 3a and 3b are now compatible. The vertical line is extended.

p. 25 – Figure 4 and Figure 5. As a color-blind reader I cannot distinguish between the colors used for the 26-Apr and 4-Sep flight data. Could the 26-Apr color be changed to a different color (e.g. blue)?

Yes, we have updated the figures with colour blind friendly palettes, and tested them with the following simulator:

<http://www.color-blindness.com/coblis-color-blindness-simulator/>

p. 27 – Table 6. Please define the acronyms VPDB and VSMOW.

The definitions have been added in the revised version, and now it is table 7.

Reference: Membrive, O., Crevoisier, C., Sweeney, C., Danis, F., Hertzog, A., Engel, A., Bönnisch, H., and Picon, L.: AirCore-HR: a high-resolution column sampling to enhance the vertical description of CH₄ and CO₂, Atmos. Meas. Tech., 10, 2163-2181, <https://doi.org/10.5194/amt-10-2163-2017>, 2017.