

Response to comments submitted by Reviewer #2

Dear Reviewers, Dear Editor,

Thank you very much for your detailed, encouraging and constructive feedbacks on our manuscript titled “**Radiocarbon analysis of stratospheric CO₂ retrieved from AirCore sampling**”. We have revised our manuscript based on the comments and suggestions made by the reviewers. Through this letter we specifically address (Authors Comments (AC) in normal fonts) to your comments and question (shown in *italic fonts*).

Thanking you,

Sincerely,

Dipayan Paul, Huilin Chen, Henk A. Been, Rigel Kivi, Harro A. J. Meijer

Specific comments:

Abstract: The abstract is more like an introduction than an information-dense presentation of the main findings of the manuscript. It is on the one hand too detailed, and yet at the same time important details about the technique and data reported are missing.

AC: The reviewer is right. We fully renewed the abstract.

Introduction: The introductory paragraphs seem unfocused and imprecise language is used (CO₂ is not well-mixed, for example, and amounts and specific numbers are not given when they could be – e.g., “sizes... are just enough... to perform quantitative radiocarbon measurements with relatively good altitude resolution:” just enough? Relatively good resolution? Numbers are needed to be more specific; there is no reason to be so vague on so many details...)... Primary references for CO₂ are more appropriate than Hartmann et al. Also: Kanu et al (2016) present new measurements of ¹⁴CO₂ (published after this manuscript was submitted but that can now be included).

AC: This section has been rephrased. The reference Kanu et al (2016) has now been included.

Sampling (section 2.1): Chen et al and Mrozek et al are in preparation, and yet this leaves primary details (and figures or pictures) missing in this manuscript. More details are needed here (even if they appear in other manuscripts) to better document the whole set of techniques and apparatuses used for the work presented here and to evaluate it.

AC: The reference Mrozek et al (2016), which is the most relevant, is now available online. The other reference Chen et al is in preparation. We deem a detailed description of the AirCore sampling beyond the scope of this article, but we have added some more details.

What does “natural” level of ¹⁴CO₂ mean for the reference gas? When was the reference gas produced? “Natural” is always changing, isn’t it? It is unfortunate it has not been measured yet by the authors or

others, but at least being more precise here could help convince readers that this is under control (perhaps). Also, why not use lower altitude AirCore air to try to compare with tropospheric values? Sampling problems and contamination issues could be more directly compared and evaluated. In other words, why only the stratosphere, when it's the gradient across the tropopause that is also really interesting and would give a better sense of "how good" the small size sample measurements are?

AC: Indeed, the natural level of radiocarbon is not constant. The fill air cylinder was prepared on 27th June 2013 at Sodankylä. Due to the pristine location of the sampling site, it is likely that the fill air contains modern CO₂, not contaminated with fossil CO₂. The radiocarbon content in this fill gas will be measured in future. However, as now described in the paper, our rough estimate of 20± 20 ‰ is good enough for our purpose.

Thanks a lot for the suggestion on including samples from lower altitudes. It indeed will help to compare these samples with the tropospheric values and thus enable a better quality control. We will certainly consider this addition in the future campaigns.

A schematic of the SAS would be helpful, not just referring to the Mrozek et al "in prep" manuscript. Not including important details nor an overall schematic (or multiple ones), combined with the organization and language, make the entire procedure, from AirCore flight to 14C measurement, difficult to follow as a reader and reviewer.

AC: A schematic has been included in Figure 2.

A more coherent description of how the AirCore data are 'corrected' for residual fill air originally in the AirCore tubing is needed, not just citing the Chen et al manuscript in preparation.

AC: We have added a short description in the revised version "Fortunately, the impact on the samples from the highest altitude can be accurately corrected by making use of the CO measurements that have been performed in parallel to CO₂ and CH₄, and the fact that the fill gas contains about hundredfold as much CO (7972 ± 5 ppb) than the natural air. In that way the measured CO is a direct measure for the fill gas fraction. The correction takes advantage of the fact that the CO concentration of stratospheric air is low (~15 ppb or ~0.2% of the CO concentration of the fill air). We derive a good approximate of the percentage of the fill air and the stratospheric air based on the measured CO concentration of the mixture of the stratospheric air and the fill air."

Why is it necessary to claim (again) that CO₂ is well-mixed? It clearly is not, and seasonal and annual variations in CO₂ concentrations are used by other groups to better understand and quantify strat-trop exchange and stratospheric mixing and transport. Just because the variations are small relative to the total abundance of CO₂, the more relevant parameter is the precision of the measurements relative to the variations. In this sense, CO₂ is not well-mixed, which is clear from Figure 1b. The authors lose credibility by saying CO₂ is well-mixed, when in my view there is no point to claiming that. If there is a reason to claim this, it should be explained.

AC: This part has been rephrased.

2.2 Extraction: More precise language and numbers would be useful: e.g., "near 100%" (can some actual numbers and standard deviations be given?); is "dummy" a typical usage in the authors' field for

a blank or control experimental run or flask (I haven't seen it and it seems both pejorative and not very informative whereas blanks or controls are more typically used and have specific meanings.).

AC: The extraction efficiency is better than 99.5%. The term “dummy” has been replaced with “control” sampler throughout the manuscript.

The discussion of CH₄ “variability” is unclear. And “relative” to what?

AC: This part has been rephrased.

Attribution of the CH₄ variability to production by friction of moving parts in stainless steel seems hard to believe as discussed. In the Higaki et al (2006) paper, a stainless steel canister filled with 1800 stainless steel balls was shaken for 15 min for a total of 3600 strokes to produce 3×10^{-5} mmol of CH₄. This represents 19 ppm in a 15 kPa He atmosphere, but would be much smaller, in ppm, at atmospheric pressure. Even if this is of the correct order of magnitude to affect measurements (1-2 ppm), it is unlikely that the valves are opened and closed for an equivalent of 3600 strokes or have as much interacting surface area as in the Higaki experiment. Perhaps there is another larger source of uncertainty in the methane measurements, either due to mixing during transfer to the Stratospheric Air Sampler or due to methane contamination as part of the fill gas. Fig 5a,b shows that the highest altitude samples are contaminated with fill gas to which a correction factor (for CO₂ and CH₄) is applied, for example. Other possibilities also seem more plausible than the one presented. If the authors think this is really a viable explanation, they should provide a more convincing argument about the similarities to the Higaki conditions.

AC: Although might seem hard to believe, this can be easily demonstrated by connecting a stainless steel valve or a stainless steel quick-connect (QC) to a Picarro CRDS that can measure methane. Upon turning the valve or connecting the QC, a spike of methane is always observed. Had this process been a degassing phenomenon, upon repetitive use the “apparent production” would have decreased and ultimately ceased. But this is not the case. If dry nitrogen is continuously flushed through a valve connected to a CRDS, we observe that the production of methane does not cease nor tend to decrease with repetitive turning of the valve. The amount of methane produced also depends on how fast/slow the valve was turned. We do not know the exact process that leads to the production of methane, but it may be linked to the carbon content in the steel. Higaki et al is the only publication, to the best of our knowledge, which investigated this phenomenon.

2.3 Graphitization The writing is unclear and English usage unusual, making this difficult to follow. Minor point: what's the pressure transducer used (make, model, accuracy, precision)?

AC: We have rephrased few things in this section, but it was not clear what exactly was unusual.

The measurement precisions are poor and the deviation from reference values represent large corrections needed. This is rather alarming and needs to be written about very clearly and precisely. The assumption that the contamination corrections will be the same for the stratospheric samples versus the reference materials needs to be better supported to be believable (and/or a better discussion later as to how this correction does or does not affect any quantitative or qualitative interpretation of the data points needs to be included later when the stratospheric data points are discussed).

AC: Contamination correction for small samples in the field of radiocarbon is well established. The rationale behind such a method of correction is that the reference material also undergoes identical preparation steps as does the samples, and thus accumulating similar levels of contamination. The extent of deviation of the reference material from its consensus value provides a measure of the contamination acquired during the preparation, which is also used to correct the sample material.

In addition, there is a discussion about “correcting the ^{14}C in CO_2 measurements of stratospheric air samples and CO_2 from reference air.” Even with this correction for modern and dead carbon contamination, there are several outliers in the standards measured. The ANU standard measured about 100‰ too low, the GS51 Standards are all measuring low (by up to 75‰, and the Ref air seems to have a mass dependence. Perhaps, these low measurements of standards, which were prepared in a similar way to the actual samples, can help explain the very low measurements from the AirCore samples (Flt. 16-07-2014). It seems that more work could be done to ensure that the reproducibility of these small-volume measurements is sufficient.

The counting statistics (leading to a 1 sigma precision of +/- 30 per mil) should be mentioned or discussed here.

In addition to the data and the spread shown in 4 for the various reference gas runs, the averages and standard deviations (as well as propagation of errors from including the 30 per mil counting statistics precision) should be given somewhere in the text or a table, and discussed.

AC: The graphitization and AMS analysis section has been extended to give the (activity-dependent) size of the corrections. Also the agreement or lack of it, of the various reference gases is now discussed in somewhat more detail. Furthermore, some sentence is added to explain that of the $\pm 30\%$ final uncertainty the largest part is still due to counting statistics. If at all, we would argue that not reference air #2, but rather GS51 has a mass dependence, with slight overcorrection for the larger masses (which would then be logical).

Compensating for small graphene targets: Is there evidence that the AMS does not sample the bottom layer of the target (i.e., that only the actual sample is being sputtered), and that having a non-uniform target (graphite + iron on top, pure iron pellet on bottom) does not affect the measurements in any way? Do the various reference gas measurements inform this?

AC: The clean iron pellet behind the sample pellet does not affect the measurements at all. The graphite containing iron pellets are not completely consumed in one 40 min measurement and can typically last for two 40 min measurements. Only after the sample has been completely sputtered, the clean iron pellet placed behind the sample pellet gets exposed, which is indicated by zero ion current for the C12 and the C13 ion beams in the AMS.

3. Results: Lines 20-25 are redundant with previous sections of the manuscript. Why not just show results and start the discussion them?

AC: Changed

I cannot follow the lines of argument in Lines 28 (pg 7) to Line 20 (pg 8), yet they are critical to understanding some of the possible shortcomings of the sampling and measurements now as well as how

they might be improved in the future. It is critical that this be intelligible to the general reader who is interested in the technique and judging its strengths and limitations.

AC: We have critically reread our own text, and tried to clarify some crucial parts by adding sentences.

*~Line 30: *Production* of radiocarbon is approximately equal between the stratosphere and troposphere (e.g., Masarik and Beer); it happens to build up in the stratosphere since there are no stratospheric sinks, but the production itself is not a maximum in the stratosphere, as stated here.*

AC: Referring to Figure 7 of Masarik and Beer: At higher latitudes, e.g., at Sodankylä (67.4° N, 26.6° E), the production of radiocarbon is almost entirely (>85%) in the stratosphere.

Figure 6 (data treatment and discussion): The $\Delta 14C$ results from the two flights are quite different (negative values for $\Delta 14C$ in flight 16-07-2014 due to “contamination in SAS sampling or [the] extraction process”). For this reason, the results would be better discussed separately rather than lumping them together. While the average of SAS 2, 3, and 4 for both flights may be close, the large negative values for SAS 1,2 and 3 for flight 16-07-2014 should discredit it from the discussion, as is noted later in the paragraph. Simply taking an average of all the lower altitude stratospheric points, which include hugely negative $\Delta 14C$ values as well as relatively large positive $\Delta 14C$ values that are ~100 per mil apart, and concluding that they ‘match’ what is expected from the troposphere at about 20 per mil, is not satisfactory. Moreover, the $\Delta 14C$ of the lower stratosphere is reported to be $10 \pm 8\%$ in the abstract (1-4) and $8 \pm 8\%$ in the results (SAS 2-4) both of which are said to represent the current tropospheric value of 20‰. The uncertainty of the measurements due to counting statistics alone (as the authors state) is much larger than this reported error ($\pm 30\%$) and then averaging together the very different values from the 2 flights should also, at least by visual inspection, should give a larger uncertainty than ± 8 per mil. Consistency in reporting is needed (or an explanation of why the 30 per mil error is shown in the figure but is not included in the reported averaged values in the abstract and here due to propagation of error).

AC: This section has been entirely rephrased. We thank the reviewer for his/her critical remarks specifically on these points

The correction for the highest altitude sample apparently depends on an estimate that the fill gas CO₂ is 20‰. Depending on the source of the CO₂, this could be inaccurate. This should be discussed in more detail.

AC: This section has been rephrased. Indeed by not knowing the radiocarbon content in the fill gas, correction can be inaccurate. Since the fill gas contamination is ~11%, the resulting correction is small. However, for future campaigns, determination of radiocarbon in fill gas will be performed.

Also, the vertical profiles presented in Fig 6 do not show the uncertainty in the altitude (vertical resolution) for the samples. What altitude resolution is expected for these samples?

Suggestion: Add sample numbers to the plot, consistent with the numbers discussed in the text of the paper.

AC: Sample number added. The estimated vertical resolution of AirCore profiles in the stratosphere is ~4 mbar based on 4-hour's molecular diffusion. As each section of the SAS contains ~50 ml, corresponding to a sampling range in pressure of 36 mbar, the uncertainty in the altitude for the ¹⁴C samples is therefore mainly determined by the sampling range in pressure. We added the estimated altitude uncertainty

Line 30 (page 8) to Line 10 (Page 9): It is unclear what contribution these lines make to interpretation of the data; the logical flow of the topics is disjoint and not clear, and the language describing strat-trop exchange ("STE") needs to be edited by a colleague who is expert in this topic (as it is not clear as written whether there is an English usage problem or lack of knowledge about STE). And what is the relevance of the discussion of bomb 14C? It seems gratuitous here, and the language later in this section seems to imply that it is STE and not fossil fuel 14CO2 propagating up into the stratosphere that has caused the 'drop' to the 'present value of ~20 per mil, which is a strange thing to imply unless one is discussing and comparing specifically with stratospheric 14CO2 measurements from the 1960s and 1970s (which is not the case). The lines and discussion following this are also disjoint, make claims that are difficult to support, and have odd atmospheric usage (e.g., the 'troposphere-stratosphere boundary' is usually referred to as the "tropopause").

AC: We came to the conclusion that the present quality of the data does not justify a discussion about stratospheric phenomena yet. For this to happen, both precision and accuracy need to be improved (as we discuss in the outlook part). Therefore, we have removed this part of the discussion.

*Top of page 10: This is where the argument is made that these are 'proof of concept' measurements "showing that the AirCore sampling for radiocarbon determination in stratospheric air is feasible. As the sampling method is relatively cheap, regular [should be defined] sample collections are affordable, leading to better understanding of the 14C budget and STE transport [sic] mechanisms involved." This remains to be seen. The language here and throughout should be modified to not make such bold claims. This is what the authors hope for, but they are not there yet. Thus, such claims cannot yet be supported. ***Again, I do not object to 'proof of concept' being published – as long as overly strong claims are not made, as long as the work is meticulously documented in clear and precise language, and as long as a clear exposition of 'next steps' to be taken to make solid improvements is given. The manuscript as written does not yet meet these standards. ****

AC: We have generally rephrased our discussion into more "modesty", but we have also tried to quantify why we think this technique can be improved such that it will be useful. In general, we have also tried to be (even) more detailed in our explanations.

Conclusions: Much of what is here should be in the discussion, not conclusions. The conclusions should summarize the manuscript, not bring up new issues to understand the new techniques and to discuss future work needed to improve the measurements.

AC: We have renamed the chapter conclusions and outlook, as we think our conclusions are closely connected to the prospects of the technique.

An additional point, re: Validation of AirCore sampling for isotope measurements While Chen et al (in prep) may provide more details on the validation and reproducibility of AirCore profiles in Finland, a brief discussion about the validity of using AirCore samples for isotope work is needed (work already

done or work planned for the future). For example, Karion et al, 2010, validate their work with comparisons between the Air-Core measurements, traditional flask measurements, and aircraft measurements for CO₂ and CH₄ mixing ratios. They also discuss the altitude resolution of the AirCore in detail, as it differs depending on molecular mass and storage times. For radiocarbon, what magnitude of fractionation should one expect in the AirCore sampling, beyond the risk of fractionation during transfer to the Stratospheric Air Sampler or to the graphitization process? What could be expected – that might be different for isotope compositions than for concentrations, if anything. This should at least be speculated on in this manuscript.

AC: Fortunately, of all the things that complicate ¹⁴C measurements, fractionation is not one of them. Isotope fractionation effects are always corrected out of the signal using ¹³C, and that happens also here. Nevertheless, the method as a whole should be validated carefully. The suggestion by the referee to deliberately also sample tropospheric air is probably the simplest way of getting a reliable intercomparison point.