

Interactive comment on “Radiocarbon analysis of stratospheric CO₂ retrieved from AirCore sampling.” by Dipayan Paul et al.

Anonymous Referee #2

Received and published: 28 April 2016

Paul et al. describe a new effort to collect, process, and analyze air from an AirCore apparatus flown by balloon into the stratosphere for measurements of the radiocarbon content of stratospheric CO₂. Unlike previous atmospheric CO₂ radiocarbon measurements, which require ~1 liter of whole air (STP) to provide adequate CO₂ to graphitize and run on an accelerator mass spectrometer (AMS), the samples collected with the AirCore (after partitioning and storage into 6 discrete samples in a “Stratospheric Air Sampler” right after the AirCore flight) are ~20 times smaller. These smaller samples required development of a “small-volume high-efficiency” graphitization procedure, which is also described here, in addition to development of an apparatus to cryogenically extract CO₂ from the 6 discrete air samples in the “Stratospheric Air Sampler” (SAS) taken from the AirCore. The 9 stratospheric air samples measured for ¹⁴C presented here are a ‘proof-of-concept’ demonstration that very small air samples col-

C1

lected by the AirCore technique can in principle be measured for ¹⁴C. However, the precision just from counting statistics alone (stated by the authors as +/- 30 per mil) is poor relative to variations of interest in the atmosphere, and there are clearly, as the authors also discuss, some large contamination issues which appear to be as large as -70 per mil – not unexpected problems given the small sample sizes. It is unclear from the manuscript as written how these problems will be overcome to provide scientifically useful data.

As a reviewer, while I do not object to publication and documentation of the heroic attempts to measure ¹⁴C on small samples (since these attempts may move the field forward even though they are not yet producing scientifically useful data), there are major flaws in the interpretation of the data and its shortcomings, as well in the clarity of presentation and even proper use of atmospheric terminology and scientific context. Furthermore, while this is a ‘technique’ paper, details of several major components of the analysis – e.g., the SAS apparatus, how to ‘correct’ profiles for fill gas at high altitudes in the stratosphere – that affect how we can understand and evaluate the measurement technique and its current limitations – are left to ‘manuscripts in preparation’ and thus leaves this technique paper rather incomplete, as this manuscript stands by itself for now in review. Overall, if one is to publish ‘proof-of-concept’ data that fall short at this point in time of being scientifically useful, there should at least be a thoughtful, thorough, and accurate accounting and assessment of the techniques developed and used and the meaning of the results to date, which are currently lacking in this manuscript as written. These major flaws in (1) lack of sufficient details presented here (which should not just be in ‘in preparation’ manuscripts), (2) logical flow and clarity of presentation, and (3) interpretation of very noisy data for both the stratospheric samples and the small reference gas analyses need to be addressed.

Some specific comments:

Abstract:

C2

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.

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).

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.

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?

C3

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.

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.

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.

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.).

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

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

C4

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.

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)?

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).

In addition, there is a discussion about "correcting the ¹⁴C in CO₂ measurements of stratospheric air samples and CO₂ 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

C5

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.

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?

3. Results:

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

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.

~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.

C6

Figure 6 (data treatment and discussion): The $\Delta^{14}\text{C}$ results from the two flights are quite different (negative values for $\Delta^{14}\text{C}$ 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 $\Delta^{14}\text{C}$ values as well as relatively large positive $\Delta^{14}\text{C}$ 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 $\Delta^{14}\text{C}$ 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).

The correction for the highest altitude sample apparently depends on an estimate that the fill gas CO_2 is 20‰. Depending on the source of the CO_2 , this could be inaccurate. This should be discussed in more detail.

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.

Line 30 (page 8) to Line 10 (Page 9): It is unclear what contribution these lines make

C7

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 ^{14}C ? It seems gratuitous here, and the language later in this section seems to imply that it is STE and not fossil fuel $^{14}\text{CO}_2$ 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 $^{14}\text{CO}_2$ 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”).

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 ^{14}C 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. ***

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.

An additional point, re: Validation of AirCore sampling for isotope measurements While

C8

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 AirCore 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.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2015-377, 2016.