Response to RC1

We sincerely thank Reviewer 1 for the very detailed and constructive review. Below we provide a point-by-point response to all comments. Line numbers refer to the revised manuscript.

Sauze et al., (2025) developed a method for the preparation and handling of small air sampling vessels with a volume of 1 mL STP for the analysis of stable carbon isotopes in CO_2. The authors describe a series of experiments that lead to a proposed protocol for the preparation and handling of the air sample containers. They find the preparation steps are critical to achieve their best precision. The authors suggest this method opens a door to new analyses of stable carbon isotopes in CO_2 from environments where the amount of available sample volume is critically limited, such as rhizosphere and chamber studies. Those applications hold the potential for important improvements of our understanding of carbon cycle processes, which is of great importance to atmospheric science and biogeochemistry, and the future of life on Earth as we know it in general.

I share the view that sampling volumes that small for high precision analyses is technically challenging. The topic of the manuscript is well suited for publication in AMT. However, I'd suggest significant rewriting, including further analysis existing data, and potentially making additional measurements. Some of the results do not seem to be of sufficient quality. The manuscript lacks fundamental basics on conventions and guidelines in isotope research. I will list a few examples and refrain from commenting on details.

We thank the reviewer for recognising the technical challenge of achieving $\delta^{13}C$–CO_2 analysis from only 1 mL of air. We acknowledge that the manuscript required clarification and restructuring to improve transparency, understanding, and reproducibility in other laboratories. In response, we have added methodological details where needed.

Given that several comments from the three reviewers highlighted similar concerns, we have also substantially revised the manuscript to more clearly emphasise the core motivation and originality of our approach: developing a simple, low-cost, and rapid method based on standard, commercially available equipment, specifically tailored for analysing very small air volumes. These points were not sufficiently explicit in the original version, and the revised manuscript now consistently underlines this rationale throughout.

Accuracy

The elephant in the room seems to be the lack of accuracy in the presented experiments. The authors focus entirely on "precision" as a quality objective (I can't remember seeing a definition of precision, but assume standard deviation of a set of experiment results?) Almost all figures, including all of the results underpinning the finally suggested protocol (Figure 7) show significant offsets between target values and the achieved average values of a series of experiments. To me, the lack of accuracy suggests that something is not quite right with the method, and a critical experimental process not sufficiently controlled. The manuscript does not seek to explore the causes for the inaccuracy. I may be convinced otherwise, but the lack of technical details leaves a lot of room for speculation on the lack of accuracy. I suggest the accuracy problem to be fully explored, ideally with additional measurements, and the details to be provided.

We thank the reviewer for this important comment. In the revised manuscript, we now explicitly define the term precision (line 83) and clarify why we focus on this parameter. While repeatability is classically

defined under strictly identical analytical conditions (Belouafa et al., 2017; Squara et al., 2020), we use precision because, although the analytical setup and protocol were constant, multiple operators contributed across analytical sequences. Precision therefore more accurately reflects the conditions under which the measurements were performed.

Regarding accuracy, we agree that the apparent offsets between target and measured values merit clarification. In the initial version of the manuscript, the uncertainties associated with our secondary standards were not explicitly shown. After including these uncertainties in figures 3-7 of the revised version, the apparent offsets are substantially reduced, indicating that the accuracy of the method is better that initially perceived. We additionally performed statistical tests comparing the certified reference value (treated as fixed secondary working standards) to the mean $\delta^{13}C\text{-}CO_2$ values obtained for each batch and the outcomes of these tests are now discuss in the corresponding sections. Finally, we provide an overall interpretation of how storage temperature significantly improved accuracy, while septum type or dual sealing did not yield any convincing benefits in this regard.

While there is likely room for further optimisation to improve both accuracy and precision simultaneously, we emphasise that, within the constraints of the method (small sample volume, minimal preparation, routine equipment), the achieved performance meets the requirements for studies in ecology and ecosystem science research, which we now better contextualise (section 4.4). This clarification, together with the quantitative assessment of standard uncertainties and the statistical comparison described above, has been added to the revised manuscript.

Protocol and gases used to prepare sample vials

The authors report a difference when flushing the vials for 8 s with CO2-free air. Unfortunately, the flow rate of that flushing is not stated, which should be reported to understand the protocol. Afterwards, the protocol includes four cycles of evacuation to 0.1 bar, followed by filling with pure N2. The authors find a substantial reduction in variability of d13C-CO2 values when the initial flush with CO2-free air is included in the protocol (Figure 2). This suggests to me that the residual CO2 is not sufficiently removed by the four evacuation-flush cycles alone. CO2-free air and pure N2 should lead to the same result, if both have sufficiently low CO2 blank, and if both gases were used to quantitatively remove residual CO2. Have the authors tested the effect when pure N2 or CO2-free air are used interchangeably for flushing or evacuation/fill cycles under the same conditions? Demonstrating the effectiveness of the sample vessel preparation with increasing initial flush flow rate and/or flush time, as well as numbers of evacuation-fill cycles would be useful to determine a protocol that leads to accurate and precise values.

We thank the reviewer for this valuable comment. We agree and included the flushing flow rate in the revised manuscript (line 212). Each 5.9 mL vial was flushed with dry CO_2-free air at 13.5 L min^{-1}, corresponding to ~40 vial-volume renewals per second. This very high flow ensures rapid and complete replacement of the internal atmosphere. For comparison, Steur et al. (2023) flushed 102 × 1 L flasks for 30 min at 3 L min^{-1}, i.e. about 90 volume renewals over 30 min, more than two orders of magnitude lower than in our setup. Flushing alone, however, is not sufficient to fully condition the vials. The subsequent four N_2 evacuation–refill cycles ensure near-total CO_2 removal and leave the vials filled with pure N_2, which minimise nitrous oxide formation in the IRMS ion source.

We did not test flushing with N_2 instead of CO_2-free air, as the presence of O_2 during the refill phase could favour nitrous oxide formation. Tests with He and Ar were also not satisfactory due to diffusion or practicality issues and cost. Blank tests performed under higher sensitive peak-detection settings

(minimum peak height = 5 mV; start/end thresholds = 1 mV s^{-1}) revealed small peaks of 10–15 mV in roughly one-third of the blanks (i.e. no peak in the other ones under those settings). However, this background remains well below the CO_2 load of actual samples and contributes negligible variability relative to our ± 0.1 ‰ precision target. These results are now reported in the revised manuscript (section 2.1).

Standard practice in atmospheric science

It should be noted that preparing vessels for atmospheric sampling and isotope analysis by evacuating, flushing and even thermal treatment is absolute standard practice. A comprehensive protocol for glass flasks was previously published (Steur et al., 2023, DOI:10.1080/10256016.2023.2234594). Successful protocols work based on a large number of gas exchanges within the vessel to eliminate remains from previous samples (memory) as well as dealing with the small but significant amount of surface water on the internal surfaces.

We thank the reviewer for this remark and agree that evacuation and flushing procedures are well established in isotope analyses, as outlined in Steur et al. (2023). Our protocol follows the same principles for moisture removal and leak prevention but is adapted for a different context: the preparation of 1 mL atmospheric samples in septum vials, where aliquot-based approaches used for large flasks (≥ 1 L) are not applicable. We have clarified this distinction and the motivation for a simple, rapid, low-cost and standard material-based protocol in the revised manuscript (line 74).

Use of d18O-CO2 as indicator of analytical performance

The protocol described by Steur et al., 2023 is especially relevant for the analysis of oxygen isotopes in CO2, which are not considered in this manuscript. d18O-CO2 can be a very useful indicator for analytical problems. Therefore, I wonder if the d18O-CO2 data could help to identify the cause of the inaccuracy? For the purpose of method refinement, the precision (standard deviation) of d18O-CO2 from different experiments might be useful.

We thank the reviewer for this helpful suggestion. We agree that $\delta^{18}O–CO_2$ can serve as a sensitive diagnostic parameter for detecting analytical issues such as leakage, fractionation or incomplete equilibration, and that its inclusion could further strengthen the evaluation of method performance. In the present study, however, the reference CO_2 used for calibration was not assigned a certified $\delta^{18}O$ value, which prevented reliable interpretation of $\delta^{18}O–CO_2$ results.

We added a statement in the section 4.4 acknowledging the potential of $\delta^{18}O–CO_2$ as an additional quality indicator and outlined its integration as a planned improvement in future developments of the method (line 420).

Control of residual CO2 and possible impact on d13C

The authors evacuate their sample vessel to 0.1 bar. In other words, around 10 % of the previous gas would still be present in every following preparation cycle. This seems to include the internal volume of the manifold (Figure 1), which is substantial in comparison to the volume of individual sample vials. This could potentially result in different gas compositions across the vials, especially as the manifold is filled with N2 from one side, pushing the gas from the previous filling towards the other side (pump side) of the manifold, where the vials at the pump end may potentially receive larger fractions of the previous gas filling. This contains some degree of speculation on my side, but I am not convinced that four cycles of evacuating to 0.1 bar and filling with N2 are a guarantee for quantitative replacement of the previous gas. It only takes 1 % of atmospheric CO2 with a d13C of around –8 ‰ and 99 % of a

working standard CO2 with a d13C of around –36 ‰ to cause an offset of 0.3 ‰. Even a 1 % blank of a small sample peak might appear relatively small in the blank test the authors performed (line 93). Because of the small sample volume, the presented method is targeting, system banks are very important, which the authors are well aware of. Vessels with septa can easily be evacuated to fractions of a mbar. Why did the authors choose not to evacuate to much lower pressure levels?

We thank the reviewer for spotting this error. The reported pressure was indeed incorrect — vials were evacuated to 0.02 mbar, not 0.1 bar. At this pressure, the residual gas in a 5.9 mL vial is negligible (~0.0001 mL). Repeating the evacuation–refill cycle four times ensures near-complete removal of the previous gas through geometric dilution. We corrected this value and clarified that, under these conditions, residual CO_2 is effectively eliminated (line 106).

Blank test data

The authors performed blank tests (i.e., line 93), but do not present blank data, instead stating they didn't find a detectable blank signal. I haven't yet seen a system that has virtually no blank. There has always been some blank, and that blank is ideally smaller than the defined limits, above which Isodat/Qtegra automatically identifies a peak. Just because the software does not report the peak, this doesn't necessarily mean there is no blank. Especially when measuring isotopes in small sample quantities, a small blank can have significant impact. I am not totally convinced that the variability in the "no flush" scenario shown in Figure 2 is not resulting from incomplete removal of the previous gas in the vial (memory). The d13C range in the "no flush" experiment is almost 2 ‰. A quick back of the envelope calculation suggests that around 6 % of ambient air CO2 with d13C of –8 ‰ would be needed to shift a CO2 with d13C of –38.7 ‰ by 2 ‰. This amount might well be detectable in the peak sizes of the measurements. It doesn't seem that insufficient memory and blank control are fully explored in the manuscript and the underlying experiments. These information or experiments should be delivered in a manuscript that seeks to establish a new sample vial preparation protocol as the primary objective.

We thank the reviewer for this comment. We are not entirely sure we fully understand the point being raised, but we note that the variability observed in the "no flush" scenario (Figure 3) indeed reflects the effect of incomplete removal of residual gas. This is precisely why the initial dry CO_2-free air flush step is essential in our protocol. After this step, followed by the N_2 evacuation–refill cycles, blank levels are negligible, and the vials are ready to receive the sample without introducing measurable isotopic bias.

Few words were also added about memory effects (line 171). In the experiment described by Siegwart et al., 2023) highly contrasting samples were analysed in random order—ranging from very high CO_2 concentrations (10,000–20,000 ppm) with enriched signatures (up to 3.5 AT%) to atmospheric levels (~420 ppm) with depleted values (~ –10‰). No measurable carry-over was observed, indicating that memory effects are negligible in our setup and require no additional correction. This is now clarified in the revised manuscript.

For peak detection, standard analytical settings typically use a minimum peak height of 50 mV with start/stop thresholds of 20 and 40 mV s^{-1}. To more rigorously assess blank integrity, we reprocessed all blank chromatograms using much more stringent criteria, lowering the minimum peak height to 5 mV and the start/stop thresholds to 1 mV s^{-1}. Under these conditions, small CO_2 peaks (10–15 mV) were detected in about one-third of the blanks (i.e. no peak in the other ones under those settings). This low-level background is well below normal detection limits and contributes only minimal additional variability, remaining negligible compared with the analytical uncertainty of the $\delta^{13}C$ measurements. These detection settings and their rationale are now described in the revised manuscript (line 116).

Undisclosed modifications to the IRMS instrument

The authors state their IRMS method includes modifications from Fiebig et al, (2005) to work for ambient CO2 mixing ratios (line 57), and "adapted here for high precision at trace levels of CO2" (line 112). However, there is no reference, no description or proof of what is done differently from Fiebig et al., (2005) and how that improved the analysis. I regard this as essential information. What are "trace-level CO2" in an atmospheric research journal? A fraction of lower tropospheric mole fraction averages?

We thank the reviewer for pointing out the lack of clarity regarding the modifications introduced relative to the method of Fiebig et al. (2005). In the revised manuscript, we now explicitly describe how our protocol differs from the original carbonate-based setup and why these adjustments are required for analysing ambient air samples with very low CO_2 content (line 146).

The new text clarifies:
(i) that our method targets atmospheric CO_2 at ~420 ppm, corresponding to ~0.2 µg C per analysis, i.e. an order of magnitude less carbon than in Fiebig et al. (2005);
(ii) that our approach analyses CO_2 directly in the ambient air matrix rather than in a He-flushed environment;
(iii) the chromatographic modification (PoraPlot Q column operated at 35 °C instead of 70 °C) to ensure better separation at low CO_2 levels;

Insufficient specification of used components

A large part of the success of the suggested sample vessel preparation protocol seems associated with the use of Terostat. Terostat seems to be a brand name for a range of sealing products and not a unique product. At no place do the authors explain what specific product they use and what it is made of, so it is impossible for a reader to gauge, or even to follow and adopt. Also, the description of how this is applied to top, and bottom could be more detailed, as the authors suggest this is a significant part to achieving high data quality.

We thank the reviewer for this comment. The term "Terostat" was indeed inappropriate and has been replaced with "butyl-rubber compound". We now provide a clear description of the material used: Teroson® RB 81, a malleable self-adhesive butyl-rubber compound with excellent resistance to gas and moisture transfer. In our basic protocol, the compound is first applied over the entire surface of the septum and vial cap (section 2.2). For dual-sealing, it is additionally applied around the lower part of the cap—directly beneath the sealing ridge and near the thread area—to reinforce protection against micro-leakage (section 3.4). These details have been clarified in the revised manuscript (line 137 & 270) to ensure full transparency and reproducibility of the procedure.

Accurate description of system performance in context of literature

The authors state that the method is of "high" precision in title and throughout the text (i.e., line 113). However, 0.1 ‰ precision in an air sample of 1 mL STP is not particularly high or novel, i.e., Brand et al, (2016), DOI:10.1002/rcm.7587, achieve 0.04 ‰ for d13C in CO2 on 1 mL or air with GC-IRMS. Schmidt et al, (2011), https://doi.org/10.5194/amt-4-1445-2011, achieve 0.05 ‰ on around 3 mL pre-industrial air sublimated from an ice core sample. These methods have demonstrated accuracy to within their much smaller measurement uncertainty or better. The additional challenge and potentially the source of the additional uncertainty/inaccuracy of the method described by Sauze et al., (2025) may thus be associated with the sample vials, not really with the sample amount.

We thank the reviewer for this remark. We acknowledge that the precision reported by Brand et al. (2016) and Schmidt et al. (2011) is higher than ours, but the analytical conditions are not directly comparable:

- Brand et al. (2016) rely on an aliquot-based approach: a large gas volume (1–5 L) is sampled once, and multiple 1 mL injections are taken from the same reservoir, allowing repeated measurements and statistical averaging. In contrast, our method analyses the entire 1 mL sample in a single injection, meaning no re-measurement or averaging is possible.
- Schmidt et al. (2011) work under much higher signal intensities, reaching ~30 V on m/z 44, whereas our analyses operate at ~2–3 V. The resulting signal-to-noise ratio inherently limits achievable precision. In addition, without accounting for sample preparation, Schmidt et al. analyse 7 samples in 25 h, while our setup allows ~20 samples to be analysed within a standard working day.

Our objective differs fundamentally from these studies and was not (enough) clearly stated in the original manuscript: provide a practical compromise that enables high-precision $\delta^{13}C–CO_2$ analysis using basic equipment, with short turnaround time and minimal sample volume, in a way that can be easily reproduced in other laboratories. We clarified this distinction in the revised manuscript (see introduction & discussion).

Data quality objectives and indicators for instrument performance

The authors seem to develop their quality control criterion around the precision value of 0.1 ‰. At no point do they provide an explanation why this value is needed for any analytical purpose. The smaller the precision, the better the protocol seems to be the paradigm. Given the precision criterion is the only data quality objective the authors seem to apply, the rationale for the choice of this value should be presented. A good example to question that approach is shown in Figure 6: The values from storage temperatures of –20C show a precision of 0.24 ‰ and are distributed around the target value with good accuracy. In contrast a precision of 0.1 ‰ is found at –80C, but then the values appear inaccurate. Yet, the storage at –80C is preferred because of the better precision, accepting inaccurate results. I'd suggest being very cautious of that rationale. Accuracy and reproducibility are at least as important as precision.

We thank the reviewer for this important and constructive comment. We agree that the choice of a 0.1 ‰ precision target must be justified. Our rationale for using 0.1 ‰ as a benchmark is both practical and methodological. In the atmospheric and ecosystem tracer community, a precision of ~0.1 ‰ on $\delta^{13}C–CO_2$ is widely regarded as the upper threshold for high-quality measurements in chamber-based, Keeling plot, and soil–plant exchange studies, where isotopic variations of 0.3–1 ‰ typically reflect biological or mixing processes (Pataki et al., 2003; Tu et al., 2001; Joos et al., 2008; Breecker et al., 2014; Leitner et al., 2023). Achieving 0.1 ‰ precision therefore provides sufficient sensitivity to resolve biologically meaningful changes without imposing excessive analytical complexity or sample volume.

Modern IRMS and field-capable systems, including Gas Bench II configurations, routinely achieve < 0.2 ‰ external precision for $\delta^{13}C–CO_2$ from 12 mL vials (manufacturer specification; Giammanco et al., 2017; van Geldern et al., 2014 for field deployments; instrument/metrology development programmes aiming at ~0.1 ‰ uncertainty). Our objective was to reach comparable analytical performance —0.1 ‰ repeatability— but with one order of magnitude less sample gas (1 mL instead of 10–12 mL). This reflects a deliberate balance between analytical resolution, operational simplicity, and experimental constraints such as chamber headspace volume and sampling frequency.

We clarified this rationale in the revised manuscript and explicitly justify the 0.1 ‰ target in the introduction (line 90).

Composition of applied gases and gas equipment

Unfortunately, the authors do not disclose the compositions of the applied gases. Besides "ambient" there is no statement on the CO_2 mole fractions in any of the applied gases. In a technical manuscript on CO_2 isotope analysis, with particular focus on small sample sizes, knowing CO_2 mole fractions of the applied gases is essential to understand experimental processes and results. Including data on the composition of used gases is obligatory for such a manuscript. For all experiments, the compositions of the gases and especially the CO_2 mole fractions must be stated. The manuscript should be clear on mole fraction scales, impurities, calibration uncertainties, etc., as well as manufacturers and models of pressure regulators on those cylinders.

We agree with the reviewer that the compositions and specifications of all gases used should be explicitly reported. This information has been added in the revised manuscript for the three working standards (WS1, WS2, and WS3; line 177). All working standards were synthetic air mixtures supplied by Air Liquide ("mélange Crystal" France) containing either 380 or 450 ppm CO_2, 20 % O_2, with the remainder N_2. They are calibrated against secondary standards, which in turn are referenced to NOAA primary standards on the RAMCES platform (LSCE, Gif-sur-Yvette, France), with assigned values of 380.0 ± 0.9 ppm CO_2 and -8.42 ± 0.10 ‰ $\delta^{13}C–CO_2$. The "ambient air" cylinder from Air Liquide was verified against the same reference scale and is used with an HBS 200-3-2.5 pressure regulator.

Isotope conventions

The manuscript ignores basic isotope conventions. All isotope reference gases used need to be stated with uncertainty and traceability chain (Camin et al., 2025, https://doi.org/10.1002/rcm.10018). This is important best practice, even though this manuscript does not show atmospheric data that a reader can compare to other measurements. The authors may have used a cylinder containing liquid CO_2 as one of the reference gases (Figure 3), and possibly for some of the gas mixing etc. It should be stated whether this contains liquid CO_2 as well as gaseous CO_2, as this may affect the isotopic composition over time or with different use, i.e., when used for mixing. When referring to an isotope ratio, the isotope ($d^{13}C$) is combined with the molecule (CO_2) as $d^{13}C$-CO_2 or $d^{13}C(CO_2)$ when referred to isotope values, where the "delta" is italicised. Negative isotope values are expressed with a long dash, rather than simple dash (Coplen 2011, https://doi.org/10.1002/rcm.5129)

We thank the reviewer for highlighting this point. In accordance with best practice (Coplen 2011; Camin et al., 2025), and as mentioned before, we now report all reference gases (WS1, 2 & 3) with their full specifications, uncertainties, and traceability (line 177). No liquid CO_2 cylinders were used at any point, ensuring that phase-change fractionation cannot occur.

We also corrected all isotope notation: δ is now italicised, isotope ratios are consistently written as $\delta^{13}C$-CO_2, and negative values use the appropriate en-dash. These corrections bring the manuscript fully in line with current isotope-reporting conventions.

References

The manuscript includes a lot of old references. There is nothing wrong with old references and credit should be given to original ideas, but in many cases, things have moved on and improved over several decades.

We thank the reviewer for this remark, we have updated the manuscript by adding several recent and relevant references throughout the revised manuscript.