Response to RC1

We are grateful to the reviewer for carefully reassessing the revised manuscript and for providing these additional comments. All line numbers mentioned in our responses refer to the revised version of the manuscript.

General comments:
The authors made significant improvements to the manuscript. In my view, there remain four areas of concern, which limit the quality of this work.

1) The language still appears to be overstating in some parts of the manuscript, when toning it down would achieve more. It is well understood that it is challenging to develop and maintain a robust analytical framework to measure isotopes in CO2 in small volume samples. It is beneficial to be open and transparent with remaining challenges and imperfections.

We thank the reviewer for this comment and agree that careful and transparent presentation is essential for small-volume isotope analyses. The manuscript has been revised throughout to moderate potentially overstated language and adopt a more cautious tone. Claims of robustness or broad applicability have been reformulated to reflect that performance was achieved under controlled conditions. We have also clarified remaining limitations, particularly regarding storage duration, material effects, and the need for further validation.

2) The use of language amounts to misleading information when the small, yet existing offset between the target values of the applied WS and the experiments are ignored as indistinguishable. The authors themselves acknowledge these offsets in a following discussion, which is received as a contradiction of earlier statements by the reader.

We thank the reviewer for this comment. We agree that earlier wording could have been interpreted as suggesting that small offsets between working standard targets and measurements were negligible. In the revised manuscript, these offsets are now transparently presented in both the Results and the Discussion, with statements clarified to indicate that values remain within analytical precision (± 0.1 ‰) and that small systematic deviations are discussed in the context of precision and potential fractionation. This ensures consistency and transparency throughout.

3) Measurement details: It is great that the authors provide a few more details on the composition of the gases. However, the provided information falls short of the information isotope papers are required to provide (i.e., Camin et al., 2025), which was requested in the initial comments. The traceability of the measurements is not clear, the authors state that they use NOAA values. However, to my best knowledge, NOAA doesn't measure isotopes, these are measured at INSTAAR in all NOAA samples and reference gases NOAA provides. Are the isotope values from NOAA/INSTAAR used to assign the final isotope values to the samples? What is the nature of the pure CO2 for the square peaks in Figure 2B? Does this cylinder contain liquid CO2 and how is this calibrated? Is this gas used for value assignment of the final samples? What is the traceability chain, the reference material these measurements are based on? What is the estimated uncertainty/bias resulting from the WS gases having a composition of 20% O2 in N2, other than a natural air matrix (78% N2, 21% O2, 1% Ar)? What is the applied 17O correction? The measurements potentially

cover a wide isotopic range from NBS19 and the atmospheric range at INSTAAR, to the range of –35‰ of the WS. How is scale compression accounted for?

We thank the reviewer for this detailed and important comment. We have revised the manuscript accordingly.

First, we now explicitly state that we are end-users of calibrated gas standards. The working standards (WS) used in this study were calibrated in-house against two secondary reference gases analysed and certified by the ICOS-RAMCES platform (LSCE, Gif-sur-Yvette, France). Additional details on calibration procedures, reference materials, and scale realisation at ICOS-RAMCES have been added to clearly document the traceability chain to VPDB (line 188).

Second, we clarified the role of the rectangular CO_2 peaks shown in Figure 2B. These peaks were generated from compressed high-purity CO_2 and serve exclusively as an internal reference for peak positioning and signal normalisation within each analytical run. They were not used for isotopic value assignment. Final δ^{13}C-CO_2 values were calculated relative to VPDB using the in-house calibrated WS (line 163).

Regarding gas matrix composition, because all calibrations and measurements were performed using GC-based separation coupled to continuous-flow IRMS (rather than optical techniques), and because standards and samples were analysed under identical conditions, no significant matrix effect on δ^{13}C-CO_2 is expected. This is now stated explicitly in the revised manuscript (line 190).

Concerning the isotopic range, the study focuses on a relatively narrow interval close to the working standards. We did not span the full VPDB scale; therefore, no large-scale compression correction was required within this restricted range. This limitation is now clearly acknowledged (line 198). For applications covering a broader isotopic range, a multi-point calibration using custom-prepared cylinders (e.g. 0 to –40 ‰) subsequently certified by ICOS-RAMCES or another recognised calibration laboratory would indeed be more appropriate to ensure scale linearity over an extended range.

Finally, regarding the applied ^{17}O correction, δ^{13}C values were calculated using the correction implemented in the Isodat software, which follows the formulation of Santrock et al. (1985) (https://doi.org/10.1021/ac00284a060) to account for the contribution of $^{12}C^{17}O^{16}O$ to the m/z 45 signal based on measured oxygen isotope ratios. According to information provided by a Senior Technical Specialist at Thermo Fisher Scientific, this corresponds to the standard ^{17}O correction implemented in the Isodat data-processing routine.

4) The discussion section does not fulfil the common goal of a discussion section, where data should be discussed to demonstrate the performance and the shortcomings of this method. Only after that could it potentially seem appropriate to list potential applications of the method and future work. In my view, at least half of the text in the discussion sections would be better placed in the conclusions. Therefore, the discussions and the conclusions sections should be re-structured or re-written.

We thank the reviewer for this constructive comment. The Discussion and Conclusion sections have therefore been substantially restructured. The revised Discussion is now organised into three focused subsections: (i) analytical performance (precision, blanks, comparison with conventional volumes), (ii) mechanisms and drivers of isotopic drift during storage, and (iii) clearly defined operational limits and methodological constraints (e.g., storage duration, precision loss, absence of δ^{18}O validation, need for inter-laboratory testing). Applications are now only briefly introduced after these limitations are explicitly stated. The Conclusion has been shortened and rewritten to provide a concise synthesis of capabilities, practical limits, and future developments, without repeating detailed discussion points.

We believe this restructuring clarifies the performance assessment of the method and better aligns the manuscript with standard expectations for Discussion and Conclusion sections.

Specific comments:
L15: …using a malleable… to minimise gas leakage… (it doesn't totally prevent leakage or diffusion, just makes it less significant and thereby workable over practical time scales).
We agree with the reviewer's suggestion and now use "minimise" instead of "prevent" leakage (line 16)

L17: "Standard material" is misleading, can equally refer to "measurement standards" or "commonly used materials", or both. It would be useful if this was more precise.
Indeed, in the context of our article, it can be misunderstood. We replaced "standard material" with "widely available materials" in the manuscript (line 16).

L22: What are low impact d13C-CO2 measurements?
The wording has been revised to explicitly indicate minimal perturbation of the sampled system, and the ambiguous term has been removed (line 23).

L23: Here and in other places in the manuscript: The authors still appear to try hard to maximise impact, when toning it down might be more helpful to allow the reader to make up their mind based on the presented facts. This point has been made by the other reviewers as well and should be followed. This writing style risks to achieve the opposite effect of what the authors intend to achieve here. "Toning it down" feels appropriate for the degree of novelty of this paper and is strongly recommended. After only reading the first paragraphs of the revised manuscript, it feels appropriate to suggest going through the manuscript again to do that. (I might be misled by a few incidences at the early part of the manuscript). The word "accessibility" is listed twice, all studies are constrained by time and costs. I'd suggest stopping that sentence after "…constrained by sample volume."
Modified (line 24) and throughout the manuscript following the reviewer recommendation.

L44: The statement "remain unsuitable" is only true for off the shelf instruments, and it should be mentioned appropriately. Laser-based instruments have been developed that measure single air samples of 1 mL for d13C-CO2 with precisions of 0.04 ‰, published 5 years ago (https://doi.org/10.5194/amt-13-6391-2020). Again, the writing style creates the feeling that the paper could be improved by toning it down.
The sentence has been revised to reflect the reviewer's comment and we are now also citing the indicated article (line 46).

L65: Here and everywhere else: the term "concentrations" should be avoided and mole fractions used instead (as done in L108).
Corrected throughout the manuscript (lines 51, 57, 65, 145, 147, 179, 180, 257, 260).

L65: "maintaining sufficiently high analytical performance", or "while achieving our data quality objective for measurement precision of 0.1 ‰."
Modified as "…while achieving our target measurement precision of ± 0.1 ‰ on δ^{13}C-CO_2." (line 66).

L87: This is quite repetitive; the analytical precision achieved with this method was mentioned only 20 lines before.
Sentence removed.

L91: Change ecologicak to ecological
Modified (line 85).

L122: The blank analysis is a great addition! Looking at Figure 2B, it seems the m/z44 amplitude is around 2V. If this was a typical sample measurement, this could be used to state that the blank peak of max. 15 mV would result in a blank contribution of less than ~1%, which would be a valuable statement here?
A sentence has been added to clarify that the blank contributes less than ~1 % of the sample signal (line 116).

L136: Please specify what "a simple septum vial" is, or change the sentence to something like "injected through the septum into the target vial."
Modified with the later suggestion (line 133).

L137: Thanks a lot for adding this sentence. It is a lot easier to understand and follow.
Thank you for the suggestion.

L138: Can any gas leakage really be effectively prevented? Or would "to minimise gas leakage" be more accurate?
The term 'prevent' has been replaced by 'minimise' on line 137, with similar changes applied consistently throughout the manuscript where appropriate.

L148: I am confused. This paragraph begins with the reference to Figure 2A, which shows that a needle with He flow is inserted into the sample vials to flush the sample through the Nafion and into the cryotrap, thereby removing N2, CH4, most of the O2, etc. The purified sample is then taken through a GC column with a He carrier gas, followed by a second Nafion before injection into the open split/IRMS. Thus, the CO2 from the air sample is separated and carried to the IRMS in a He stream for d13C analysis. This seems consistent with the description of the protocol in L154-L168. However, the authors say here that the analysis is directly performed on the atmospheric gas matrix, rather than flushing the sample vial with He etc., which seems to me is exactly what is shown in Figure 2A. I might be getting this wrong, but I feel it is important to clarify this as it seems very contradictory.
We thank the reviewer for pointing out this potential source of confusion. We have revised the paragraph to clarify that, although helium is used as the carrier gas within the Gas Bench system in both approaches, our vials are initially pre-filled with nitrogen in which the sampled 1 mL air to be analysed is inserted. The sample (pre-filled N + 1ml sampled air) is then transported by a helium carrier to the IRMS. In contrast, Fiebig et al. (2005) used vials pre-saturated with helium containing CO_2 generated from acid–carbonate reactions. The revised text now clearly distinguishes between the initial vial gas matrix and the downstream helium carrier flow, removing the apparent contradiction (line 148).

L151: I am not sure I understand what “optimise separation peak shape” means. Lowering the temperature of the GC column increases peak separation, but to my understanding, will reduce the sharpness of the peaks. Please clarify.

The text has been revised to clarify that lowering the GC column temperature represents a compromise that improves separation from other atmospheric components while maintaining adequate peak definition and stable retention times (line 155).

L160: These rectangular peaks are typically generated from a cylinder with liquid CO2. The authors have stated in their response that no cylinder with liquid CO2 is used in this system. Please include this statement here, especially as it is mentioned that the isotopic composition of this gas is somehow calibrated to the VPDB scale. It is of course OK for this gas to be in the cylinder in the liquid phase, but it needs to be very clearly stated that this gas is only used as mediator and not for final value assignment of the samples. Alternatively, if it is used for final value assignment, this needs to be expanded on, to include how the isotopic composition was assigned to the VPDB scale, and what materials were used etc. Follow Camin et al., (2025), which the authors refer to in their response.

There is a misunderstanding here. The five rectangular reference peaks visible on the figure 2B are effectively generated from pure CO_2 supplied from a cylinder, which necessarily contains a liquid phase. As this is standard practice in IRMS operation, we did not originally provide further details on this gas, as it is indeed used solely as a reference signal and mediator for peak positioning and signal normalisation, and not for final isotopic value assignment.

The calibration gases used for δ^{13}C-CO_2 value assignment and explicitly described in the manuscript (WS1, WS2, and WS3) are atmospheric air mixtures and do not contain a liquid CO_2 phase. Our previous statement referred exclusively to these working standards and to the gases used for their calibration to the VPDB scale. We have added details to explicitly clarify this distinction in the manuscript and avoid further confusion (line 163).

L169: It seems to me that most GC-IRMS systems produce a single peak per sample injection, so this statement is confusing. The wording that this is in contrast to “traditional” systems seems misleading, as there is no novelty with a single injection per sample to be claimed. Rather, it confirms the feeling that this may be fishing for novelty in an attempt to make this manuscript seem stronger and the method more novel than what they really are. There is no loss in saying that “This configuration produces a single, well-defined CO2 peak for isotopic analysis from each sample.”

Corrected as suggested by the reviewer (line 176).

L171-175: Informative addition, very nice!

Thank you for this positive feedback.

L176: Please be more specific and follow Brand 2014 and Camin et al., 2025 in the description of the traceability chain. This is not clear to me. It has to include the fundamental RM this is based on (NBS19, …) it’s assumed isotope value and date. Later on, it seems that isotope values are assigned to the samples based on NOAA standards. As far as I understand, this would be INSTAAR for CO2 isotopes, which has recently migrated from their original scale to JRAS-06, with implications on the isotopic composition of dependent gases. It is important to be absolutely clear.

All δ^{13}C-CO_2 values are traceable to the VPDB scale via secondary reference gases analysed and certified by the ICOS-RAMCES platform (LSCE). At ICOS-RAMCES, isotopic compositions are determined by IRMS and normalised to VPDB using international reference materials anchored to NBS19. Our laboratory acts strictly as an end-user of these certified reference gases.

We clarified that NOAA standards are used for CO_2 mole fraction calibration at ICOS-RAMCES, whereas isotopic values are assigned within the IRMS calibration framework and not directly by NOAA. The revised text now explicitly distinguishes between mole fraction and isotopic traceability to avoid ambiguity (line 188).

L178: Here and everywhere else, please ensure consistent use of – and - in d13C–CO2 or d13C-CO2 throughout the manuscript.

Homogenised throughout the manuscript.

L180: The reference gases have no Argon and only 20% O2, which is different from a natural air matrix with 21% O2 and 1% Ar. This may cause a significant offset, especially if these gases were calibrated using optical instruments for mole fractions or isotope ratios. The authors should clarify to what degree the calibration of these gases and the dependent measurements may suffer from this effect. It should be stated if there is no effect to be expected and a reason for that be provided. Otherwise, an estimate for the degree of uncertainty resulting from this effect should be provided in the discussion of the analyses.

The absence of argon and the slightly lower O_2 content in the synthetic air standards were not expected to significantly affect $\delta^{13}C\text{-}CO_2$, as all calibrations and measurements were performed using IRMS and GC-based techniques rather than optical instruments, and samples and standards were analysed under identical conditions. This has been now explicitly addressed and clarified in the revised manuscript (line 193).

L182: RAMCES platform needs to be explained or referenced. What does this mean? What type of instrument is used for that? Laser-based? Can the magnitude of this effect on the sample results be estimated that stems from the difference in the matrix of the working standards WS1-WS3 and the matrix of natural air?

Few words were added about the ICOS-RAMCES group, including the analytical techniques used for CO_2 mole fraction and isotope calibration, and explicitly described the traceability chain to NOAA primary standards. We also added a short justification explaining why the slight matrix differences between synthetic air working standards and natural air are expected to have a negligible impact on $\delta^{13}C\text{–}CO_2$ at the achieved precision (line 189).

L192: Figure 2B would be more relevant if it included a zoom to the baseline or the 45/44 ratio trace, to show that separation and stability are achieved, as stated in L237.

Figure 2B has been revised to include a zoomed view of the baseline (lower panel), allowing clearer visualisation of signal stability and chromatographic separation, as described in line 255. This addition demonstrates the flat baseline between injections and the absence of co-elution affecting the CO_2 peak.

L232: The red area in Figure 3 can either indicate the measurement precision, or the uncertainty of the d13C-CO2 value for WS1, but not the precision of WS1.

We have clarified the wording by using “measurement precision of WS1-3” as suggested (lines 249, 277, 299, 329 and 349) to accurately reflect that the shaded area represents the standard deviation derived from repeated measurements of WS1-3.

L235: Not sure if “described” is the right word here? Could this be changed to “…is shown in Figure 2B”?
Modified as suggested (line 253).

L237: See above comment on L192, these statements could be demonstrated in Figure 2B.
We agree and have revised Figure 2B accordingly. A zoomed view of the baseline has been added to visually demonstrate chromatographic separation and signal stability, as stated in line 255.

L263: What bias is this referring to? The lack of accuracy resulting from storage at room temperature? More clarity on this is needed here.
The term “bias” has been clarified to explicitly refer to the storage-induced isotopic drift observed at room temperature (line 281).

L265: Can “while” be changed for “because”? Or can this sentence be broken in two: “These tests suggests that the septum configuration is not the cause of the 1‰ offset. Additional sealing and sample preservation strategies must therefore be explored to improve the stability of small-volume air samples.”
The sentence has been rephrased and split into two, as suggested (line 283).

L279: What is “simple sealing” in Figure 5? Using the butyl-rubber at the septum or the threaded end of the cap only, in comparison to using it at the septum and the threaded end of the cap? I also assume simple sealing cannot refer to the septum+cap on the vial without any butyl-rubber, as this would be the same as shown in Figure 4, and should lead to a 1‰ offset? Please be more specific and relate the terminology in the text to the terminology used in the figures.
The three sealing configurations are now clearly defined in Section 2.2: (i) no additional sealing (septum and cap only), (ii) simple sealing (butyl-rubber applied to the septum and cap surface), and (iii) dual sealing (butyl-rubber applied to both the top and lower threaded part of the cap) (line 133). Each configuration is now explicitly referred to in the relevant sections (lines 267 and 287) to ensure consistency with the figures.

L287: “…at room temperature.” It seems the word temperature is missing.
Corrected (line 305).

L303: Please add a sentence on the small offset between WS2 target and the test results in the -80C data shown in Figure 6.
The offset is mentioned line 319 as follows: “… mean δ^{13}C values exhibited a very slight positive offset relative to their respective working-standard targets. However, as this offset remains within the margin of the combined analytical precision for both standards and samples, the measured values are sufficiently robust for comparative analysis, where the relative trends are the primary focus rather than absolute accuracy.”

L334: I am not fully convinced that these are fully indistinguishable. There seems to be a consistent pattern in Figure 6 and 7 suggesting that the d13C-CO2 averages are shifted towards 13C enrichment by about 0.1 ‰. I strongly feel that this cannot be ignored and has to be mentioned. The authors themselves start to discuss this in L397, confirming this suspicion. However, this acknowledgement comes out of the blue and is a stark contradiction to the claims of "indistinguishability" the authors communicated throughout the manuscript.
We qualified our statement to note that a very slight positive offset is observed after short-term storage, which remains small relative to the combined analytical uncertainty of the standards and samples. The discussion has also been nuanced and include the evolution of both precision and accuracy over time (line 356).

L340: Similar to comment on L17, please be more clear when using terminology such as "standard material", when referring to commonly available materials.
The Discussion and Conclusion sections have been substantially revised. Where appropriate, the term "standard material" has been replaced with "widely available commercial materials" (line 430 and consistently throughout the manuscript) to avoid ambiguity.

L343: What is "straightforward" vial preparation? Can "straightforward" just be deleted?
Deleted.

L340: I'd suggest using the discussion section to discuss experimental results and potential limitations or unknowns. Section 4 sounds like a sales pitch or a conclusion section.
Section 4 has been substantially rewritten to focus explicitly on experimental results, performance assessment, and methodological limitations (line 365).

L378: I am not fully convinced that the presented method "effectively prevents isotopic drift", but that it "effectively reduces isotopic drift" to enable high quality measurements when following the described protocol. The authors state this themselves in section 4.3, when they recognise that the offset patterns were sometimes bidirectional, hence most of the time they were not. Therefore, please tone the statements of indistinguishability etc down, they add no value, instead, the opposite might rather be the case.
As suggested, we rewrote section 4 (line 365).

L405: A systematic rationale also expressed in this sentence could be included in the introduction: The smaller the measured sample volume and the lower the CO2 mole fraction, the stronger the impact of systematic analytical errors on the results. As this method seeks to optimise a method for analysis of very small analyte amounts, it is of paramount importance to identify and eliminate all factors that contribute bias and uncertainty, including those that seem irrelevant in the analysis of large sample volumes.
A sentence expressing this systematic rationale has been added to the Introduction (line 38).

L419: It was suggested that the authors might want to consider showing their d18O-CO2 data. However, they decided not to do, providing the lack of a reliable d18O-CO2 calibration as the reason to not show those data. This is unexpected, as INSTAAR, the laboratory that probably provided the isotope calibration scale, definitely measures d18O-CO2 very well. However, even without a solid

calibration, d18O-CO2 data could help to diagnose a wide range of potential problems, which the authors state themselves in L420. In the presence of an unresolved offset, is a mystery that this data opportunity is not even considered for exploitation.
We agree that δ^{18}O-CO$_2$ data could be very informative. Although the primary reference gases provide δ^{18}O-CO$_2$ values, our secondary working standards (WS1–3) were not calibrated for δ^{18}O-CO$_2$. As the raw data are no longer available, retrospective analysis is not currently possible. Under these conditions, we feel it is safer not to include any δ^{18}O-CO$_2$ results due to lack of traceable calibration.

L431: Again, this paragraph, starting with a suggestion for future work, sounds much more suited for the conclusions than the discussions.
This paragraph has been moved to the conclusion (line 430).

L440: The conclusions appear as a shorter version of the discussion section. These sections have different purposes and should therefore have a different rationale.
The Conclusion has been rewritten to avoid repetition of the Discussion and to better reflect its distinct purpose. It now provides a concise synthesis of the main findings, clearly states the operational limits of the method, and briefly outlines perspectives for future work, without re-discussing detailed results or mechanistic interpretations. We believe this revision clarifies the complementary roles of the Discussion and Conclusion sections (line 430).

L444: Standard based… Please see comments on confusion with standards above.
This misleading information has been corrected throughout the manuscript.

L456: …without significant isotopic drift…
The Conclusion has been shortened and rewritten (line 430).

Response to RC2

We thank the reviewer for re-evaluating our manuscript and providing a last comment. Line numbers in our responses refer to revised manuscript.

General comments:
I consider that the authors have made their current best efforts to improve the manuscript according to the long lists of reviewer's comments. I suggest the authors to consider a minor correction: In the captions of Figures 3–7, "the shaded red band its precision" is newly inserted, but here the use of "precision" is strange. The term "precision" usually represents capability or performance of an instrument, but here the author would like to indicate possible range of a value. Here, for instance, I would reformulate the wording to "the determined δ13C-CO2 value of WS1 and its uncertainty." But this could cause a conflict in use of the terms as the authors seem to avoid "uncertainty."

We thank the reviewer for his positive feedback and for this constructive remark.

We agree that the term "precision" must be used carefully. In this case, the shaded red band represents the standard deviation calculated from repeated measurements of WS1-3, i.e. the dispersion of replicate analyses under repeatable conditions. It therefore reflects the measurement precision of WS1-3, rather than the uncertainty of an assigned reference value.

To avoid ambiguity, and in agreement with RC1's related comment, we have revised the figure captions to read: "the measurement precision of WS1-3" (lines 249, 277, 299, 329 and 349). This wording more accurately describes what is shown (the standard deviation of a measurement series) and avoids confusion.