

Comments to “Development and evaluation of a suite of isotope reference gases for methane in air” by Sperlich et al.

Summary of manuscript

The authors set up measurement systems for anchoring $\delta^{13}\text{C-CH}_4$ and $\delta^2\text{H-CH}_4$ to the international reference isotope scales. They also prepared a suite of isotope reference gases (pure CH_4 gases and synthetic CH_4 -in-air gases) and calibrated them. The synthetic CH_4 -in-air gases will be made available for intercomparison to achieve high compatibility in the atmospheric monitoring community.

General comments

This study serves a fundamental contribution to the research community in study of the global CH_4 cycle. Increasing number of data of CH_4 isotopic ratios in air is becoming available, especially since the development of continuous-flow measurement system. However, datasets from different laboratories cannot be currently merged due to unnegligible and unidentified inter-laboratory calibration offsets. This hampers optimized use of isotope data for better understanding of the CH_4 cycle. This study is a start up of anchoring measurements of different laboratories. The authors set up the current best available measurement methods, systems and calibration strategies. Further intercomparison efforts to distribute the synthetic CH_4 -in-air standard gases developed in this study to worldwide laboratories are also an important next step. This study has significance with good measurement quality, and the manuscript is fairly well organized; however, many parts are not written in a clear and concise way and do not contain complete information. In my opinion, this study should be definitely published but the current manuscript version is not suitable for publication. I realize that this manuscript will be an important long-term reference for the research community. This is why I think that the manuscript should meet high level of transparency for all detailed information and clearness in description to help following researchers. It is regrettable that number of my comments below includes points that could have been thoroughly considered and corrected with the authors' responsibility before submission. I would like to encourage the authors to complete all details and rewrite the whole manuscript to improve the readability, after which this manuscript might be evaluated again.

It is very important to unify terms throughout the manuscript (not to mix up

different terms for same meaning). In particular, I suggest to clearly define hierarchy of the gases at early part of the manuscript and always use the terms defined. Name of the gases are bracketed by “” (e.g. “Megan”), but I do not think this necessary. I also suggest modifying Figure 1 so that readers easily find which gas is at which level (exact correspondence between Figure 1 and Tables 1 and 2). You might consider a figure similar to the attached figure and get it merged into Fig. 1 (see also specific comments to Fig. 1). This would improve traceability. To my understanding, the hierarchy is (1) IAEA “reference materials”, (2) MPI-BGC “working standards” (H₂O and carbonates), (3) “master” CH₄ gases (or “MPI-BGC primary” might be an alternative), (4) “secondary” CH₄ gases—these are also called “calibration” in tables, but I think “secondary” (or “MPI-BGC secondary”) used at other places fits better, and (5) “synthetic CH₄-in-air standard” produced by dilution of the secondary.

Since iSAAC system plays an important role, the authors should elaborate at least overview of the system (e.g. materials and temperature of preconcentration traps, separation columns, etc.). Is it similar to Brass and Röckmann (2010)? There are many references citable. The authors might briefly describe the system even if they consider a separate publication. One more point on iSAAC—it seems that iSAAC is operated with “Carina-1” with nominal values on the IMAU scale, but this is not written explicitly. The authors should clearly describe this in the manuscript.

The IMAU scale is in the manuscript considered to be an established scale independent from the MPI-BGC scale developed in this study. Description of origin of the IMAU scale is also important for perfect peace out of the manuscript. I suppose that the IMAU scales (both $\delta^{13}\text{C}$ and $\delta^2\text{H}$) are also ultimately referenced to some of the IAEA reference materials. The authors should mention to this, not only cite to the technical paper Brass and Röckmann (2010). I understand that, as appeared in this study, during the course of the long history, systematic offsets could arise due to different measurement methods, different maintenance strategies etc. even that both scales are on the identical VPDB and VSMOW scales in theory—that is, the problem is in calibration at individual laboratories. It is important to describe this fact. This also helps future intercomparison planned.

Also confusing is that the gas “Carina-1” has $\delta^{13}\text{C}$ and $\delta^2\text{H}$ values on the IMAU scale. The authors might present those values with a note that they are on the IMAU scale, since they are listed in Table 2.

Specific comments

P2 L2: “Isotope ratios of trace gases” to “Isotope ratios of CH₄”.

P2 L2: I do not think isotope ratios are still “an emerging tool”—they have tested over decades.

P2 L3: Quay et al. (1991, 1999) are also important studies that showed quantitative use of isotopic compositions for understanding the global CH₄ cycle, which could fit this context.

P2 L16: “papers” to “the paper”

P2 L17ff: The authors might list the IAEA reference materials used in the cited references. Also, the authors might mention to the fact that most (or all?) calibrations in the cited references were linked to the IAEA certified materials and that in theory on identical scales. It would highlight the fact that calibration offsets have arisen during scale propagations in individual laboratories—this probably corresponds to more ambiguous word “the diversity of referencing trajectories” (P2 L21) in the manuscript.

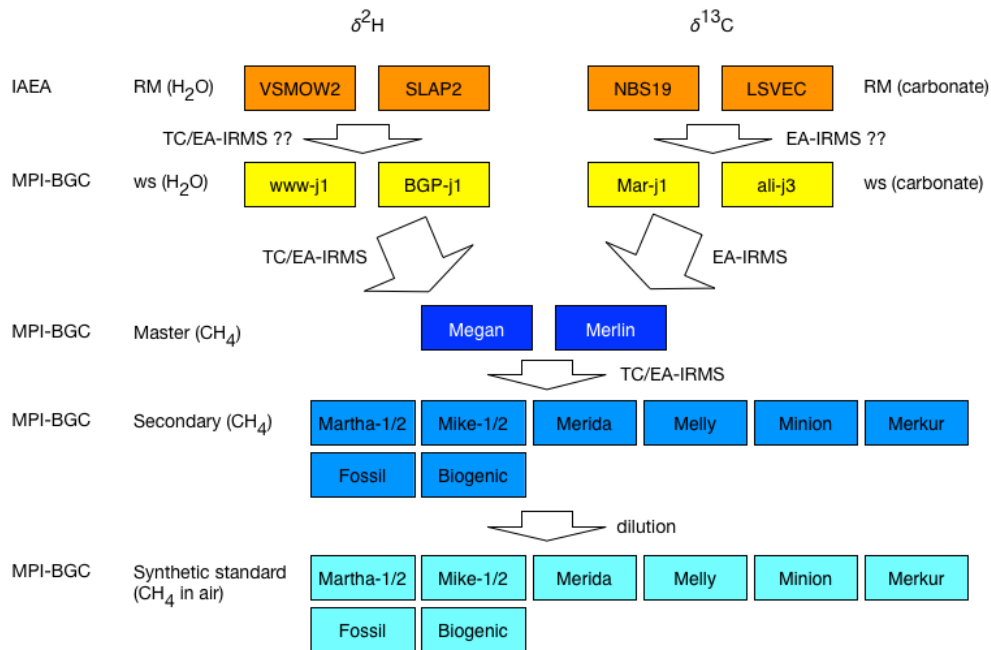
P2 L21: First, here in the context the authors should refer only to papers already published and thus “W.A Brand pers. comm.” should be left out; second, personal communication with one of contributing authors is strange—same for other places. An alternative might be “W. A. Brand, unpublished data” if this matches the context. Otherwise unpublished results should not be cited unless clear necessity—it would not help readers and traceability of the manuscript.

P2 L22: “referencing” to “calibration”

P2 L23: I suggest to start this paragraph with a phrase like “For isotope ratios of CO₂ in air, ...”, so that readers better understand that the authors intend to introduce an analogical method already applied for an other gas.

P3 L7ff: Please unify use of “the Principle of Identical Treatment” and “PIT” everywhere in the manuscript.

Figure 1: As pointed out in the general comments, the authors might enrich this figure by adding information on which standards in Table 1 are at which level in this figure. An other idea is to split this figure both for $\delta^2\text{H}$ and $\delta^{13}\text{C}$ as the figure below. Another good reference is Figure 2 of Sperlich et al. (2012).



P3 L15: I do not think that “commercial provider”, “methanogenic origin” and “isotopic origins” should be listed in parallel with equal stress. Differences in methanogenic origin and isotopic composition in CH₄ is highly linked; the former is a primary factor for the latter. Roughly speaking, commercial providers do not always care about it.

P2 L17: “and are therefore of known isotopic composition” should be left out. The “Biogenic” and “Fossil” gases are treated as unknown in this study and removing the phrase would reduce risk of misunderstanding.

Table 1: I am not sure if “create” is a suitable word. I think “produced” sounds more common.

P4 L7: The contents in the first paragraph of section 3.1 should be moved here.

P4 L15: “We show the most recent values...” The table gives references, but the caption mentions only to Brand et al. (2014)—this is confusing. If original papers are cited in the table, I would delete this sentence of the caption.

P5 L4: The abbreviation “IRMS” appears here for the first time.

P5 L6: “after the conversion to H₂”—since there is “to convert CH₄ to H₂” just a one line above, this phrase is a redundancy. This sentence may be like “...to convert CH₄ to H₂ (+carbon) for subsequent measurement of $\delta^2\text{H}$ -CH₄ in pure CH₄ gases on IRMS.”

P5 L4–L9: In my first reading, I (probably) misunderstood that CH₄ and H₂O in sample are converted to H₂ in different reactors. Besides the above suggestions, I suggest possible reorganization of the sentences: “We use a TC/EA coupled to an IRMS via an

open split for $\delta^2\text{H}$ measurements of CH_4 and water; CH_4 in sample gas is introduced into the TC/EA and converted to H_2 in a glassy carbon reactor maintained at 1450°C ; water sample is injected through a heated septum at 130°C into the same reactor and converted to CO and H_2 ; the converted H_2 is measured on the IRMS for $\delta^2\text{H}$.”

P5 L9: “hydrogen isotopic composition” to “ $\delta^2\text{H}$ ”

P5 L11: “that is configured as shown in Fig. 2” to “configured as Fig. 2”

P5 L12: “Typical CH_4 feed flow rates range between 2-3 mL/minute.” to “Typical CH_4 feed flow rate is 2–3 mL/minute.” How much volume of CH_4 is actually injected?

P5 L14: Leave out “now”. My suggestion is like: “Measurement sequences are configured by combining injections of CH_4 gas and reference water materials into the reactor, which are via the 10-port valve (Fig. 2) and septum from an autosampler, respectively.” I would leave out “, where both CH_4 and H_2O are converted...”

P5 L16: I would leave out “Apart from the injection procedure...” because it comes up later soon again.

P5 L18–19: “The amounts” to “Amounts”. “to achieve matching peak...” to “to match peak...”. “during IRMS analysis” to “of the IRMS output signal”. But this sentence might be left out because same (and more in-detail) explanation appears page 7.

P5 L20: “Fig. 4” to “Fig. 3”

P6 L8: “This configuration enables alternative injections of CH_4 gases with known and unknown $\delta^2\text{H}$ values in an identical fashion; the calibrated “Megan” and “Merlin” served as known reference gases for calibrations of the other CH_4 gases (Table 1).” I do not get what “the respective isotope scales” means.

P6 L14: “applying” to “employing”

P6 L15: “working reference waters” to “working standard” because they seem to be abbreviated as “ws” in Table 2. Decapitalize www-“J”1 and BGP-“J”1 to harmonize with Table 2.

P6 L20ff: use “PIT” throughout the manuscript if this term is abbreviated.

P7 L4: leave out “several”, otherwise write more explicitly.

P7 L8: “assumed to be” kept constant

P7 L13: “..., which guarantees same level of H_3 -factor correction and allows use of the standard integration software (ISODAT, Thermo, *Company name & place as other places*)”

P7 L14–18: “We performed...” Please write explicitly about what were tested. I do not

understand which experiments in appendix correspond to these descriptions. I could not find descriptions that support these sentences—it sounds like that the authors argue good performance of their measurements without showing any clear evidences.

P9 L5: “routinely used”—does this mean the system routinely used at MPI-BGC or at other laboratories? This sentence may intend to endorse reliability of measurements in this study, but it would not work if the authors mean the latter. Even if the former, just “routinely used” cannot justify. I would write more explicitly—for instance, “Similar systems were used for... in previous studies (...et al.)” or “MPI-BGC has operated this system for... over X years (...et al.)” The same comment is also for P5 L6.

P9 L7: My suggestion: “Outflow of the 10-port valve enters into a combustion furnace with helium carrier gas stream of 10 mL/min through an 1/16 inch tubing (70/30% Cu/Ni alloy) specially fitted to the EA system (Fig. 4).” How important is “the oxygen plume region”? If this is critical, elaborate more. Otherwise the following sentence may be enough.

Figure 4: “oxidation”—the text use different terms like “combustion chamber” and “combustion reactor”. This confuses readers. “combustion furnace” (also “reduction furnace” might be good.

P9 L13: “the combustion reactor”—is this different from “the combustion chamber” in the preceding sentence? Use one term if not.

P9 L14: “All samples are oxidized to CO_2 in the combustion furnace maintained at 1020°C (Werner et al., 1999) and experience identical analytical treatment (PIT) thereafter.” I think that the authors monitor furnace’s temperature but not temperature at which sample actually reacts. Do not mix up “combustion” and “oxidation” so often, otherwise readers wonder if the authors intend to use them with different meanings.

P9 L16: “The sample is dried by passing through a Nafion dryer (...) and a trap filled with $\text{Mg}(\text{ClO}_4)_2$, and then introduced into a GC column (3 m \times 1/4 inch Porapak PQS, CE instruments) held at 80°C .”

P10 L1–3: Circumlocution. My suggestion: “By alternating injection of CH_4 gas and carbonate reference materials such as LSVEC, Mar-j1 and ali-j3, we referenced our Master CH_4 gas (Megan and Melrin) to the VPDB isotope scale.” Later NBS-19 also appears (not listed in Table 2), and the authors describe Megan and Merlin were calibrated against NBS 19 and LSVEC. How were Mar-j1 and ali-j3 used? Were they used to assign values to the Master gases? Or just for measurement control? Clarify this.

For $\delta^2\text{H}$ calibrations, the authors used “in house standards” to assign values to the Master gases, but here for $\delta^{13}\text{C}$, the Master gases are directly measured against the IAEA materials. I do not understand what “over three independent periods each” means.

P10 L5: “analysis” to “analyses”

P10 L13–P11 L1: “The first three square-shaped peaks and the last tailing peaks are for pure CO_2 working gas and CH_4 - and Li_2CO_3 -derived CO_2 , respectively.”

P11 L2: “...(with peak widths of 101 s and...)”

P11 L7: “the two-step calibration strategy approach”—Clarify meaning of this. Does it mean the calibration has two anchoring points on the VPDB (VSMOW) scale?

P11 L8: As mentioned earlier, NBS 19 appears here for the first time and not listed in Table 2. Megan and Merlin are bracketed by “” at other places, but not here.

P11 L14: “...has been found₂ which has been...”

P11 L18: “BGC ISOLAB”—Would not “MPI-BGC” work as other places?

P11 L27: “The secondary or “calibration” CH_4 gases”—as mentioned earlier, define the hierarchy of standards early in the manuscript so that confusing words like here do not come up.

P11 L30: Which is correct “Master CH_4 gas” or “master methane gas”?

P12 L1ff: I suggest to use a term “synthetic CH_4 -in-air standard” for gases produced by dilution of the secondary CH_4 gases.

P12 L2: Same comments as P11 L18.

P12 L3: “This system (named ARAMIS) is used to produce synthetic standard gases with atmospheric CH_4 mole fraction levels.”

P12 L4: “We diluted aliquot of the secondary CH_4 gases with CH_4 -free air to...”

P12 L5–6: “The CH_4 -free matrix air, which was produced by target-mixing ultra-pure constituents, contains N_2 , O_2 , N_2O and Kr at atmospheric levels, so that composition of the produced gas is as close to ambient air as possible.”

P12 L7–8: “...to account for the interference effect on the $\delta^{13}\text{C}$ - CH_4 measurements using GC-IRMS systems (Schmitt et al., 2013).” I would leave out “so that...” but add this part to the preceding sentence.

P12 L8: “A sensitive analysis...” I do not understand what this sentence means. An upper limit for what? Sensitive to what? What is high-precision gas-chromatography? Is it different from GC-IRMS?

P12 L9: I do not think Table 1 gives “further details.”

P12 L11: “an average standard deviation” to “average standard deviations”, but what does an average standard deviation mean? Standard deviations for measurements of synthetic standard from multiple dilutions were averaged?

P12 L12: Personal communication with one of contributing authors is odd. For this sentence, the authors should give complete description of measurement precisions of the analytical system in section 2.6.

P12 L17: Same comment as P12 L12.

P12 L16–L20: As described in general comments, the authors should describe at least overview of the iSAAC system. Since it seems to be a continuous-flow system, likely similar to Brass and Röckmann (2010), the authors should give configuration of the system with basic information of key components, so that at least relevant researchers can understand the system well, because this system is a key to calibrate the synthetic CH₄-in-air standards.

P12 L22: As described in the general comments, even that readers are led to Brass and Röckmann (2010) for very details, the authors should describe basics of the IMAU scale. That is, with what types of laboratory standards they maintain long-term consistency of their scales, how and to what IAEA materials their scales are ultimately referenced. As I mentioned earlier, their scales are also the VSMOW and VPDB scales; the IMAU scales are identical to the MPI-BGC scales in theory.

P12 L23: ““Carina-1” as master reference gas for the iSAAC system” means, as long as any gases are measured by iSAAC system, they are assigned values on the IMAU scale? This should be clarified and it would help understand Table 4 and relevant texts.

P12 L24–25: Are Carina-1 and Carina-2 of identical origin and should they agree both $\delta^{13}\text{C}$ and $\delta^2\text{H}$ in theory? This is unclear. Only the description “Jena air” in Table 1 does not guarantee it. Regarding the offset in $\delta^2\text{H}$, was the cause identified? Also, the authors might present $\delta^{13}\text{C}$ and $\delta^2\text{H}$ values of Carina gases on the IMAU scales.

P12 L26: “The synthetic isotope reference gases” to “The synthetic CH₄-in-air gases”, “previously calibrated CH₄ gases” to “the secondary CH₄ gases”

P12 L27: “The results of these measurements are compared to the calibration results of the secondary CH₄ gases so that the differences between the calibrations in this study and the iSAAC measurements against Carina-1 indicate the offsets between the MPI-BGC and IMAU scales.”

P13 L4: “carbon and hydrogen isotope ratios” to “ $\delta^{13}\text{C}$ -CH₄ and $\delta^2\text{H}$ -CH₄”

P13 L7: “in the reference hierarchy”; “calibration CH₄ gas” to “secondary CH₄ gas”

P13 L8: “Therefore” to “As a result”

P13 L8: “primary reference materials” to “international reference materials”; or the authors might use “RM” for the IAEA certified reference materials, with this stated in early part of the manuscript

P13 L9: “against a Master-CH₄ gas Megan”

P13 L11: “Both Megan and Merlin are fossil in origin with typical $\delta^{13}\text{C-CH}_4$ and $\delta^2\text{H-CH}_4$ signatures (e.g. ...). The two Master-CH₄ gases are similar in $\delta^2\text{H-CH}_4$ with calibrated values of $-168.0 \pm 0.6\text{‰}$ for Megan and $-165.7 \pm 0.6\text{‰}$ for Merlin. The calibrated $\delta^{13}\text{C-CH}_4$ values are $-40.75 \pm 0.07\text{‰}$ for Megan and $-39.07 \pm 0.07\text{‰}$ for Merlin.” Megan’s $\delta^2\text{H}$ is -168.1 in Table 3, but -168.0 here in the text.

P14 L5: I would leave out lines “Fossil comparison” and “Biogenic comparison” from the table and bring them to footnote of the table. This might improve ease of viewing of the table by focusing only on calibration result of this study.

P14 L7: “were then mixed with $\delta^2\text{H}$ -spike gas to produce Martha-2 and Mike-2 with $\delta^2\text{H-CH}_4$ values higher than or similar to that of the tropospheric CH₄, respectively.”

P14 L11: “Results for secondary CH₄ gas calibrations against Master CH₄ gases”

P14 L13: “...spiked CH₄ mixtures, and thus cover wide range of $\delta^{13}\text{C-CH}_4$ (...) and $\delta\text{D-CH}_4$ (...).” I would leave out “, which include...”

P14 L15: “create” to “produce”

P14 L15: “a CH₄ gas with $\delta^2\text{H-CH}_4$ close to that of the tropospheric value, ...”

P14 L16: “a fossil CH₄ gas”—Is this the “Fossil” gas? I wonder if “diluted” is the correct word, because dilution usually means lowering mixing ratio of certain compounds, but here the CH₄ gases are almost pure gases and mixture of such gases just result in no change in CH₄ mixing ratio.

P15 L1: “Results for calibrations of synthetic CH₄-in-air standards”

P15 L2: “Aliquots of the secondary CH₄ gases were diluted with CH₄-free air to produce the synthetic CH₄-in-air standards (section 2.5) for analysis on the iSAAC system (section 2.6).” Large part of L3 is redundancy.

P15 L4: “the diluted CH₄ reference gases” to “the synthetic CH₄-in-air gases”

P15 L4: “...against Carina-1 on the IMAU scale.”

P15 L5–7: My suggestion: “We calculate the difference between the calibrations of the secondary CH₄ gases on our measurement systems (sections 2.2 and 2.3) and the

synthetic CH₄-in-air standards on iSAAC, $\delta_{iSAAC} - \delta_{sec}$ (Table 4); the value indicates calibration offsets between the MPI-BGC and IMAU scales, if we assume no isotopic fractionation in the dilution process.”

P15 L8–9: My suggestion: “Our experiments show a good agreement for $\delta^{13}\text{C-CH}_4$ with an average difference of $+0.02\pm 0.08\text{‰}$, but a significant systematic offset of $+4.0\pm 1.1\text{‰}$ for $\delta^2\text{H-CH}_4$.” What is the cause of the $\delta^2\text{H-CH}_4$ offset? The authors should discuss on possible sources of the offset here or in the next section.

P15 L13: “the pure CH₄ gas” to “the secondary CH₄ gas”

P15 L13: “A sudden drift” by what kind of reason?

P15 L19: “Comparison of the calibrations on the new MPI-BGC scale developed in this study and iSAAC measurements on the IMAU scale (Brass and Röckmann, 2010).”

P15 L21: “the name of the Master/secondary CH₄ gas that was diluted to the synthetic CH₄-in-air standard for iSAAC measurements”

P15 L23: “the mean difference” to “the average difference”—use one term both for caption and table.

P15 L24: “Therefore, this value (marked with °) was excluded for calculation of the average scale difference.”

P16 L2: “...to calibrate the pure CH₄ gases for...”

P16 L3: “methane” to “CH₄”; “...into the isotope measurement system that also analyses water and carbonate reference materials, and thus subject to PIT.”

P16 L5: “The online oxidation of CH₄ to CO₂ (and H₂O) is considered to produce no isotopic fractionation.” This sentence needs references.

P16 L6: “However, CH₄ is relatively stable chemically; complete oxidation of CH₄ thus requires high temperature and surplus of oxygen.”

P16 L7: “allow for” to “leave”

P16 L8: “CO₂ present” to “presence of CO₂”

P16 L9: “This source of analytical error” to “This effect”

P16 L10: “don’t” to “do not”

P16 L11: “In the MPI-BGC systems” to “In the MPI-BGC EA-IRMS system”; Is this also the case for iSAAC? Clarify.

P16 L12: “methane” to “CH₄” (2 places)

P16 L14: “quantitative” Does it mean combustion efficiency of 100% or close to 100% without measurable isotopic fractionation? Write explicitly. Meaning of “quantitative”

is unclear also for many other places.

P16 L15: “quantitatively”—same comment as the above. I would write: “It has been demonstrated that introduction of carbonates into high-temperature combustion furnace yields CO₂ conversion resulting in high-precision $\delta^{13}\text{C}$ measurements (...).”

P16 L16: “...oxygen isotope composition is altered completely in the conversion process from the original carbonate to the product CO₂.”

P16 L17: I do not understand what this sentence means.

P16 L18: “ambiguity” to “uncertainty”; “extracting” to “calculating”; “values” to “value”; “tends to cancel” means that it is not 100% guaranteed and there are exceptions. The authors should better justify. Besides, this paragraph seems to be readable for only expert readers who can easily refer to equations for ¹⁷O correction. The authors should present “kind” introduction at the beginning of this paragraph on why this matters—I think this needed for AMT which expect readers more general than e.g. RCM.

P16 L21: “hydrogen” to “ $\delta^2\text{H}$ measurement”

P16 L22: “quantitative conversion”—same comment as the above.

P16 L23: “methane” to “CH₄”; My suggestion: “Major artifact can arise from more variable surface adhesion of H₂O than CH₄ in the combustion furnace before they are converted to H₂ (and CO/carbon).”

P16 L24: “water” to “H₂O”; I would write: “This can lead to memory effect in the $\delta^2\text{H}$ -H₂O measurements, then corrections or discarding initial injections are needed (...).”

P16 L26: “In addition, we found a minor dependence of...”; “In the appendix” to “In Appendix A”

P16 L29: “with a large number of analyses”—number of analyses is given neither in Table 3 nor text. Without this, this description is not justified. I would write instead: “We have presented calibration results of the secondary CH₄ gases Fossil and Biogenic (Table 3). Both gases...”

P16 L30: Leave out “in an earlier study”; “the” to “a”; “combusting” to “combustion of”

P16 L31: “methane” to “CH₄”; “sampling of”; “consecutive” to “subsequent”?

P16 L31–33: “Sperlich et al. (2012) analyzed the CH₄ derived CO₂ for $\delta^{13}\text{C}$ -CH₄ on a dual inlet IRMS and the CH₄ derived H₂O for $\delta^2\text{H}$ -CH₄ on a TC/EA-IRMS system similar to this study or cavity-ring-down spectroscopy.”

P16 L33: My suggestion: “Our calibration results are in overall agreement in both $\delta^{13}\text{C-CH}_4$ and $\delta^2\text{H-CH}_4$ with the previous values by Sperlich et al. (2012) within the uncertainties of both measurements (Table 3).”

P17 L1–3: “However, our calibration results for Biogenic and Fossil appear to..., suggesting systematic offsets.”

P17 L3: This statement is strange. After this, the authors argue that calibrations by Sperlich et al. (2012) are less robust (which I do not think justified), but here the authors argue that their measurements are supported by agreement with the unreliable measurements by Sperlich et al. (2012).

P17 L5: “a large number of measurements”—same comments as P16 L29. I do not think that long-term use itself guarantees accuracy and robustness of a measurement system. Long-term use with unidentified artifacts can happen. The Kr interference on $\delta^{13}\text{C-CH}_4$ measurements is a good example—GC-IRMS had used for more than a decade until it was found.

P17 L6–8: I do not think that the statement in these sentences is justified. Sperlich et al. (2012) indeed has limited number of measurements, but did thorough treatments for complete combustion and reduction. Therefore, combustion and reduction by Sperlich et al. (2012) might be more complete than the online conversions made in this study. If so, the calibrations by Sperlich et al. (2012) might be more robust even if number of measurements is less. To keep the authors’ argument, the authors should describe weak points of Sperlich et al. (2012) specifically.

P17 L11: “uncertainty” to “uncertainties”

P17 L12: “an indicator”; “create” to “cause”

P17 L14: I do not understand what the authors argue here. With “CH₄ reference materials” (which I do not understand what the authors refer to), what would the author do for further tests? What is the authors’ best idea?

P17 L15–17: “The total propagated uncertainties in our calibrations are smaller than or similar to uncertainties of widely used analytical systems for $\delta^{13}\text{C-CH}_4$ and $\delta^2\text{H-CH}_4$ in air/ice core samples (references are needed). Therefore, a suite of standard gases developed in this study can help to increase the compatibility between international laboratories.”

P17 L21: “The number” to “Number”

P17 L22: “could lead to” to “provides”; delete “of the combined data sets”

P17 L22–23: My suggestion: “However, such merged dataset has not been achieved by the lack of reference materials that enable direct intercomparison in the community.”

P17 L25–26: The paragraph can just follow the previous one without line break; My suggestion: “To deal with this problem, we prepared 12 pure CH₄ gases (the secondary CH₄ gases) and accurately referenced them to the international isotope scales VSMOW and VPDB for $\delta^2\text{H-CH}_4$ and $\delta^{13}\text{C-CH}_4$, respectively. These secondary CH₄ gases then were diluted to produce 8 synthetic CH₄-in-air standards in 5-L glass flasks.”

P17 L26–28: I do not understand what the authors argue here. The authors say the synthetic CH₄-in-air standards were “tested for their use”, but where in the manuscript did they evaluated usability of the gases? Section 3.3 does not seem to describe this. “separately” to “indendently”.

P17 L29: “synthetic atmospheric reference gases” to “synthetic CH₄-in-air standards”; “isotopic composition of CH₄” to “ $\delta^{13}\text{C-CH}_4$ and $\delta^2\text{H-CH}_4$ ”

P17 L30: My suggestion: “These synthetic CH₄-in-air standards will help worldwide laboratories to anchor their measurement datasets to unified $\delta^{13}\text{C-CH}_4$ and $\delta^2\text{H-CH}_4$ scales shared in the atmospheric monitoring community, enabling compatible isotope ratio datasets for better understanding of the global CH₄ cycle.”

P18 L8: This appendix describes “preliminary experiments”, but the authors state in the main text that they optimized the measurement condition.

P18 L11: “This effect is avoidable by repetitive...”

P18 L13: “However” to “Moreover”

P18 L14: delete “furthermore”; “scales with isotopic difference” to “depends on difference of isotope ratio”

P18 L15: delete “only”

P18 L16: I do not understand how this sentence is linked to the preceding sentence by “Therefore”.

P18 L17: delete “for” after “corrected”

P18 L18–19: “..., as our system”—as I mentioned for P17 L5, long-term operation itself does not guarantee correctness, so the last sentence of this paragraph is not justified.

P18 L21: “We made 106 injections of an identical H₂O samples...”

P18 L22: “Septum” to “septum”; I would delete “however, there seems to...”

P18 L23: My suggestion: “A systematic increase of $\delta^2\text{H-H}_2\text{O}$ with the septum”

temperature is apparent above 90° C, but the $\delta^2\text{H-H}_2\text{O}$ value reaches the plateau around 130° C.”

P18 L24: “At three highest temperatures”—The authors say the $\delta^2\text{H-H}_2\text{O}$ value stabilized above 130°C, then the average should be calculated from the data above 130°C.

P18 L25: “The $\delta^2\text{H-H}_2\text{O}$ values stabilized above 130° C suggests quantitative adequate conversion of H₂O processing without ...”

P18 L26: “at the lower temperature range” to “below 90°C”; but my suggestion is for instance: “In contrast, the $\delta^2\text{H-H}_2\text{O}$ values below 90°C show an insignificant slight increase with the septum temperature, which deviates from the pattern above 90°C.” I do not understand the original sentence. What does “an offset” mean?

P18 L28–30: I do not understand these sentences. Elaborate better.

P19 L3: “fall onto a polynomial fit” is the authors’ interpretation. Here the authors should write, for instance, as “the black line is the quadratic polynomial fit to the data above 90°C”.

P19 L4: What is “the offset”?

P19 L6: The reasoning of taking a value from the polynomial fit is unclear.

P19 L10: “high temperature” to “high-temperature”; “utmost” to “particular”

P19 L12: I would leave out “The temperature...”

P19 L13: “...at different reactor temperatures (Fig. A2).”

P19 L16: “150 K” to “150°C”

P20 L14ff: I do not find where this appendix fits in the main text and what it supports.