

## Author's Comments to the reviewers

First of all, we would like to thank both reviewers and the author of the additional comment for their insightful and constructive critiques. As suggested by both reviewers, we have considerably re-written and re-structured large parts of the manuscript. Therefore, page and line numbers of the specific comments by the reviewers will not match the numbers in the revised manuscript. The revised manuscript considers every point mentioned by the reviewers and the comment, unless specifically addressed otherwise. We dedicated specific emphasis on the following aspects:

1) In the introduction, we present a review on the history of produced scale anchors at a range of laboratories analysing  $\delta^{13}\text{C-CH}_4$  and  $\delta^2\text{H-CH}_4$ . This list provides a good overview on the magnitude of potential inter-laboratory offsets and their variability over time. 2) We present a dedicated method section of the analytical setup that has been developed at MPI-BGC to analyse  $\delta^{13}\text{C-CH}_4$  and  $\delta^2\text{H-CH}_4$  in air samples. 3) We present a dedicated method section on the origin/development of the scale anchors at IMAU and MPI-BGC. 4) We have improved and clarified the applied terminology. For example we define the use of "calibration" in comparison to "measurement", we refer to the produced gas mixtures as synthetic CH<sub>4</sub>-in-air standards and to the pure CH<sub>4</sub> gases as primary and secondary CH<sub>4</sub> gases. "Working standard" is abbreviated WS, "certified reference material" CRM, "reference material" RM and "matrix reference material" (e.g. CH<sub>4</sub> in air) as m-RM, complying with recommendations from IAEA TecDoc 1350. 5) We revised the calculation of the uncertainties and dedicated a separate method section to present our calculation method. All data presented in the manuscript include the uncertainties of the full traceability chain where possible (CRM → WS → primary CH<sub>4</sub> → secondary CH<sub>4</sub>). The presented uncertainties include the most recent development in CRM's, i.e. the new uncertainty for LSVEC. 6) We revised the comparison between MPI-BGC and the previously published data/method of Sperlich et al., (2012). This includes a revision of the Sperlich et al., (2012) data and their uncertainties to include the full traceability chain. Moreover, we present new comparison experiments between the two methods of MPI-BGC and Sperlich et al., (2012) to discuss/support the methods presented in this manuscript. 7) The experiments to evaluate the potential for analytical errors of the new methods are explained and discussed in the main text in more detail, full details are provided in the Appendix.

Our response to the reviewers comments is indicated in blue in the following.

## Reviewer 1 (anonym)

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  $\text{CH}_4$  gases and synthetic  $\text{CH}_4$ -in-air gases) and calibrated them. The synthetic  $\text{CH}_4$ -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  $\text{CH}_4$  cycle. Increasing number of data of  $\text{CH}_4$  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 un-negligible and unidentified inter-laboratory calibration offsets. This hampers optimized use of isotope data for better understanding of the  $\text{CH}_4$  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  $\text{CH}_4$ -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.

[Thanks a lot, we hope we achieved this goal in the revised manuscript.](#)

It is very important to unify terms throughout the manuscript (not to mix up different terms for same meaning).

[Better defined throughout revised manuscript.](#)

In particular, I suggest to clearly define hierarchy of the gases at early part of the manuscript and always use the terms defined.

[OK.](#)

Name of the gases are bracketed by "" (e.g. "Megan"), but I do not think this necessary.  
[Matter of taste, we left it as it was.](#)

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.

[OK.](#)

To my understanding, the hierarchy is (1) IAEA "reference materials", (2) MPI-BGC "working standards" ( $H_2O$  and carbonates), (3) "master"  $CH_4$  gases (or "MPI-BGC primary" might be an alternative), (4) "secondary"  $CH_4$  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  $CH_4$ -in-air standard" produced by dilution of the secondary. Since iAAC 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.

[OK. Respective publication has just been published \(Brand et al., 2016\), nevertheless we dedicate an entire section to describe the measurement method in brief.](#)

One more point on iAAC—it seems that iAAC 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.

[OK.](#)

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  $\delta^{13}C$  and  $\delta^2H$ ) 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.

[OK.](#)

Also confusing is that the gas "Carina-1" has  $\delta^{13}C$  and  $\delta^2H$  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.

[OK.](#)

## Specific comments

P2 L2: “Isotope ratios of trace gases” to “Isotope ratios of CH<sub>4</sub>”.

OK.

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

“a powerful tool”

P2 L3: Quay et al. (1991, 1999) are also important studies that showed quantitative use of isotopic compositions for understanding the global CH<sub>4</sub> cycle, which could fit this context.

OK.

P2 L16: “papers” to “the paper”

“papers that ... (e.g. Levin 2012, Sapart 2013 and Schaefer 2016).”

P2 L17ff: The authors might list the IAEA reference materials used in the cited references.

OK.

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.

We list CRM’s used for scale anchoring by a range of laboratories and also include the revisions of their isotope values and their uncertainties. This is a technical but informative illustration of sources for inter-laboratory offsets. Errors due to propagation add on top of this.

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.

OK.

P2 L22: “referencing” to “calibration”

OK.

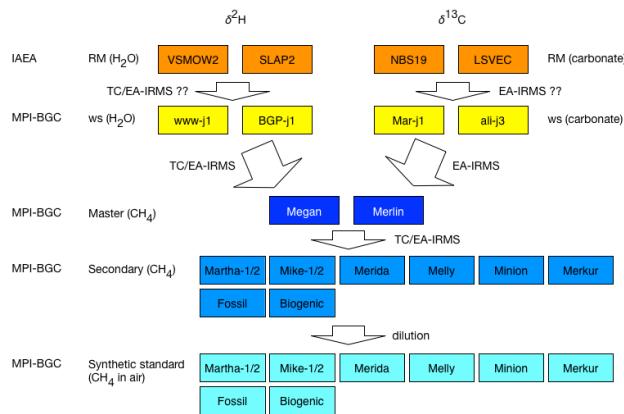
P2 L23: I suggest to start this paragraph with a phrase like “For isotope ratios of CO<sub>2</sub> in air, ...”, so that readers better understand that the authors intend to introduce an analogical method already applied for another gas.

OK.

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

OK.

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. Another 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).



OK.

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 CH<sub>4</sub> is highly linked; the former is a primary factor for the latter. Roughly speaking, commercial providers do not always care about it.

OK.

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.

OK.

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

OK.

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

OK.

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.

OK.

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

OK.

P5 L6: "after the conversion to H<sub>2</sub>"—since there is "to covert CH<sub>4</sub> to H<sub>2</sub>" just a one line above, this phrase is a redundancy. This sentence may be like "...to convert CH<sub>4</sub> to H<sub>2</sub> (+carbon) for subsequent measurement of δ<sub>2</sub>H-CH<sub>4</sub> in pure CH<sub>4</sub> gases on IRMS."

[Paragraph re-formulated.](#)

P5 L4–L9: In my first reading, I (probably) misunderstood that CH<sub>4</sub> and H<sub>2</sub>O in sample are converted to H<sub>2</sub> 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 δ<sup>2</sup>H measurements of CH<sub>4</sub> and water; CH<sub>4</sub> in sample gas is introduced into the TC/EA and converted to H<sub>2</sub> 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<sub>2</sub>; the converted H<sub>2</sub> is measured on the IRMS for δ<sup>2</sup>H."

[Paragraph re-formulated.](#)

P5 L9: "hydrogen isotopic composition" to "δ<sup>2</sup>H"

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

[Matter of taste, we prefer original wording as the plumbing of a Figure cannot be configured but the plumbing that is shown in a figure.](#)

P5 L12: "Typical CH<sub>4</sub> feed flow rates range between 2-3 mL/minute." to "Typical CH<sub>4</sub> feed flow rate is 2–3 mL/minute." How much volume of CH<sub>4</sub> is actually injected?

["Typical flow rates of CH<sub>4</sub> range between 2-3 mL/minute."](#)

P5 L14: Leave out "now". My suggestion is like: "Measurement sequences are configured by combining injections of CH<sub>4</sub> 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<sub>4</sub> and H<sub>2</sub>O are converted..."

["While CH<sub>4</sub> gases are injected manually, H<sub>2</sub>O is introduced via autosampler."](#)

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

[OK.](#)

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.

[OK.](#)

P5 L20: "Fig. 4" to "Fig. 3"

[OK.](#)

P6 L8: "This configuration enables alternative injections of CH<sub>4</sub> gases with known and unknown δ<sup>2</sup>H values in an identical fashion; the calibrated "Megan" and "Merlin" served as known reference gases for calibrations of the other CH<sub>4</sub> gases (Table 1)."

["For the calibrations of primary CH<sub>4</sub> gases, the two sample loops are fed by the same CH<sub>4</sub> gas \(connecting vent 1 and CH<sub>4</sub> port 2\). The sample loops are fed by two different gases for the calibration of secondary against primary CH<sub>4</sub> gases \(Table 1\)."](#)

I do not get what “the respective isotope scales” means.

Removed.

P6 L14: “applying” to “employing”

Section re-formulated.

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.

OK.

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

OK.

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

Re-formulated.

P7 L8: “assumed to be” kept constant

OK.

P7 L13: “..., which guarantees same level of H<sub>3</sub>-factor correction and allows use of the standard integration software (ISODAT, Thermo, *Company name & place as other places*)”

“The similarity between CH<sub>4</sub>-derived and the H<sub>2</sub>O-derived H<sub>2</sub> peaks allows the use of the standard integration software (ISODAT, Thermo Finnigan, Bremen, Germany).”

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.

OK.

1. memory

2. septum temperature

3. reactor temperature for quantitative conversion of both H<sub>2</sub>O and CH<sub>4</sub>

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.

P5 L6: OK. We strongly disagree with the speculation regarding our intention to choose the formulation. The cited references were published almost 20 years ago (1999), TC/EA-IRMS for H<sub>2</sub>O analysis is a standard method. The use of a well-established method for the production of reference materials is required by the IAEA (TecDoc 1350). We mention this explicitly in the revised manuscript to prevent from speculations regarding the intention of the formulation.

P9 L5: Same argument. EA-IRMS are workhorses in numerous- laboratories all around the world. Listing more previous studies than the two given references seems to exceed the norm on a 2-decade old technique. New is our modification that allows to use this system for gas samples.

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.

"We fitted a 1/16" tube of 70/30 % Cu/Ni alloy to the EA and used the previously described 10-port valve to inject the CH<sub>4</sub> samples into the EA with a 10 mL/min helium flow (Fig. 4)."

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.

"Figure 4: 10-port valve for manual CH<sub>4</sub> injections coupled to the EA-IRMS system through custom made gas inlet into combustion (oxidation) unit. The proportions of illustrated components is chosen to increase clarity."

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

OK.

P9 L14: "All samples are oxidized to CO<sub>2</sub> 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.

OK.

P9 L16: "The sample is dried by passing through a Nafion dryer (...) and a trap filled with Mg(ClO<sub>4</sub>)<sub>2</sub>, and then introduced into a GC column (3 m×1/4 inch Porapak PQS, CE instruments) held at 80° C."

"The sample is dried by passing through a NafionTM membrane (Perma Pure LLC, Toms River, NJ, USA, not shown in Fig 4) and a Mg(ClO<sub>4</sub>)<sub>2</sub> trap before it enters the GC column (3m, 1/4", Porapak PQS, CE instruments) held at 80°C."

P10 L1–3: Circumlocution. My suggestion: "By alternating injection of CH<sub>4</sub> gas and carbonate reference materials such as LSVEC, Mar-j1 and ali-j3, we referenced our Master CH<sub>4</sub> gas (Megan and Melrin) to the VPDB isotope scale."

Rephrased.

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.

OK.

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. **That is correct for  $\delta^2\text{H}$ .  $\delta^{13}\text{C}$  calibrations were made against LSVEC directly and one internal working standard (MAR-J1 in most cases and ALI-J1 once).**

I do not understand what “over three independent periods each” means.

**““Megan” and “Merlin” were each calibrated on three different days to determine the external reproducibility of the  $\delta^{13}\text{C}$  results.”**

P10 L5: “analysis” to “analyses”

OK.

P10 L13–P11 L1: “The first three square-shaped peaks and the last tailing peaks are for pure  $\text{CO}_2$  working gas and  $\text{CH}_4$ - and  $\text{Li}_2\text{CO}_3$ -derived  $\text{CO}_2$ , respectively.”

**Matter of taste.**

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

OK.

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?

**Due to the concerns expressed in all reviews and comments, the entire section 2.4 Measurement uncertainty and error propagation is now re-formulated.**

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.

OK.

P11 L14: “...has been found, which has been...”

OK.

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

OK.

P11 L27: “The secondary or “calibration”  $\text{CH}_4$  gases”—as mentioned earlier, define the hierarchy of standards early in the manuscript so that confusing words like here do not come up.

OK.

P11 L30: Which is correct “Master  $\text{CH}_4$  gas” or “master methane gas”?

**Now primary  $\text{CH}_4$ .**

P12 L1ff: I suggest to use a term “synthetic CH<sub>4</sub>-in-air standard” for gases produced by dilution of the secondary CH<sub>4</sub> gases.

OK. Note we produce synthetic CH<sub>4</sub>-in-air standards from both primary and secondary CH<sub>4</sub>

P12 L2: Same comments as P11 L18.

OK.

P12 L3: “This system (named ARAMIS) is used to produce synthetic standard gases with atmospheric CH<sub>4</sub> mole fraction levels.”

“The MPI-BGC operates an analytical system (named ARAMIS) to dilute pure CO<sub>2</sub> with CO<sub>2</sub>-free air to atmospheric CO<sub>2</sub> mole fraction without isotopic fractionation (Ghosh et al., 2005). We use ARAMIS to dilute an aliquot of primary or secondary CH<sub>4</sub> with CH<sub>4</sub>-free air to atmospheric CH<sub>4</sub> mole fractions (~2 ppm) in 5-L glass flasks with a final filling pressure of 1.8 bar absolute.”

P12 L4: “We diluted aliquot of the secondary CH<sub>4</sub> gases with CH<sub>4</sub>-free air to...”

We diluted and compared primary and secondary CH<sub>4</sub> gases (Table 6).

P12 L5–6: “The CH<sub>4</sub>-free matrix air, which was produced by target-mixing ultra-pure constituents, contains N<sub>2</sub>, O<sub>2</sub>, N<sub>2</sub>O and Kr at atmospheric levels, so that composition of the produced gas is as close to ambient air as possible.”

OK, see next comment below.

P12 L7–8: “...to account for the interference effect on the  $\delta^{13}\text{C}$ -CH<sub>4</sub> measurements using GC-IRMS systems (Schmitt et al., 2013).” I would leave out “so that...” but add this part to the preceding sentence.

“The CH<sub>4</sub>-free matrix air has been target-mixed from ultra-pure constituents and contains N<sub>2</sub>, O<sub>2</sub>, N<sub>2</sub>O and Kr at atmospheric levels, so that the composition of the produced CH<sub>4</sub>-in-air standards is as close to ambient air as possible. Krypton was added to this matrix air to account for the measurement artefact during GC-IRMS analysis of CH<sub>4</sub> for  $\delta^{13}\text{C}$  (Schmitt et al., 2013).”

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?

“A blank analysis of the CH<sub>4</sub>-free-air yielded a maximum CH<sub>4</sub> blank of 0.5 ppb.”

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

OK, sentence removed.

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?

OK, sentence removed.

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.

OK.

P12 L17: Same comment as P12 L12.

OK.

P12 L16–L20: As described in general comments, the authors should describe at least overview of the iAAC 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 CH<sub>4</sub>-in-air standards.

OK.

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.

OK.

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

OK, section re-worded and re-structured.

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.

OK, section re-worded and re-structured.

P12 L26: “The synthetic isotope reference gases” to “The synthetic CH<sub>4</sub>-in-air gases”,  
OK, section re-worded and re-structured.

“previously calibrated CH<sub>4</sub> gases” to “the secondary CH<sub>4</sub> gases”

OK this is clarified in the revised version.

P12 L27: “The results of these measurements are compared to the calibration results of the secondary CH<sub>4</sub> gases so that the differences between the calibrations in this study and the iAAC measurements against Carina-1 indicate the offsets between the MPI-BGC and IMAU scales.”

OK, section re-worded and re-structured.

P13 L4: “carbon and hydrogen isotope ratios” to “ $\delta^{13}\text{C-CH}_4$  and  $\delta^2\text{H-CH}_4$ ”

OK.

P13 L7: “in the reference hierarchy”; “calibration  $\text{CH}_4$  gas” to “secondary  $\text{CH}_4$  gas”

OK.

P13 L8: “Therefore” to “As a result”

OK.

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

OK., unified abbreviations throughout revised manuscript.

P13 L9: “against a Master- $\text{CH}_4$  gas Megan”

OK.

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- $\text{CH}_4$  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.

OK, section re-worded and partially shifted into section 2.1 as suggested above.

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.

OK. Revised manuscript has dedicated section for comparison between CIC and MPI-BGC.

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  $\text{CH}_4$ , respectively.”

OK.

P14 L11: “Results for secondary  $\text{CH}_4$  gas calibrations against Master  $\text{CH}_4$  gases”

OK, “Results for secondary  $\text{CH}_4$  gas calibrations against primary  $\text{CH}_4$  gases”.

P14 L13: “...spiked  $\text{CH}_4$  mixtures, and thus cover wide range of  $\delta^{13}\text{C-CH}_4$  (...) and  $\delta^2\text{H-CH}_4$  (...).” I would leave out “, which include...”

OK.

P14 L15: “create” to “produce”

OK.

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

OK, paragraph added to section 2.1 and re-worded.

P14 L16: “a fossil CH<sub>4</sub> 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 CH<sub>4</sub> gases are almost pure gases and mixture of such gases just result in no change in CH<sub>4</sub> mixing ratio.

OK, paragraph added to section 2.1 and re-worded. (diluting referred to deuterium content but that was bad choice of wording).

P15 L1: “Results for calibrations of synthetic CH<sub>4</sub>-in-air standards”

OK.

P15 L2: “Aliquots of the secondary CH<sub>4</sub> gases were diluted with CH<sub>4</sub>-free air to produce the synthetic CH<sub>4</sub>-in-air standards (section 2.5) for analysis on the iSAAC system (section 2.6).” Large part of L3 is redundancy.

OK.

P15 L4: “the diluted CH<sub>4</sub> reference gases” to “the synthetic CH<sub>4</sub>-in-air gases”

OK.

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

OK.

P15 L5–7: My suggestion: “We calculate the difference between the calibrations of the secondary CH<sub>4</sub> gases on our measurement systems (sections 2.2 and 2.3) and the synthetic CH<sub>4</sub>-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.”

Sentence now in section 2.8 “We calculate the isotopic difference ( $\delta_{iSAAC} - \delta_{pure}$ ) between the measurements on iSAAC and the calibrations of the pure CH<sub>4</sub> gases (Sect. 2.2 and 2.3), which indicates the correction to anchor the measurements at MPI-BGC to JRAS-M16.”

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.

OK.

P15 L13: “the pure CH<sub>4</sub> gas” to “the secondary CH<sub>4</sub> gas”

The entire discussion on the offset in “Biogenic” is re-worded.

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

Statement was too speculative and is removed from the revised manuscript.

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).”

Table Caption re-worded to “Table 6: Differences in  $\delta^2\text{H-CH}_4$  and  $\delta^{13}\text{C-CH}_4$  between primary/secondary CH<sub>4</sub> gas calibrations and iSAAC measurements of the synthetic CH<sub>4</sub>-in-air standards using the scale anchor based on “Carina-1”. Differences are calculated as  $\delta_{iSAAC} -$

$\delta_{\text{pure}}$ . The bottom line shows the average and the standard deviation ( $1\sigma$ ) of considered differences, excluding the value of “Biogenic” (°) as described in main text.”

P15 L21: “the name of the Master/secondary CH<sub>4</sub> gas that was diluted to the synthetic CH<sub>4</sub>-in-air standard for iAAC measurements”

[See above comment.](#)

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

[See above comment.](#)

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

[See above comment.](#)

P16 L2: “...to calibrate the pure CH<sub>4</sub> gases for...”

[OK.](#)

P16 L3: “methane” to “CH<sub>4</sub>; “...into the isotope measurement system that also analyses water and carbonate reference materials, and thus subject to PIT.”

[“Samples and reference materials were always analysed in the same analytical systems, thereby complying with the PIT as much as possible.”](#)

P16 L5: “The online oxidation of CH<sub>4</sub> to CO<sub>2</sub> (and H<sub>2</sub>O) is considered to produce no isotopic fractionation.” This sentence needs references.

[Sentence removed.](#)

P16 L6: “However, CH<sub>4</sub> is relatively stable chemically; complete oxidation of CH<sub>4</sub> thus requires high temperature and surplus of oxygen.”

[Sentence re-worded to: “Quantitative oxidation of CH<sub>4</sub> during  \$\delta^{13}\text{C}\$ -CH<sub>4</sub> analysis requires high reaction temperatures \(e.g. Dumke 1989\)”.](#)

P16 L7: “allow for” to “leave”

[Sentence re-worded to: “CH<sub>4</sub> is a potent source of protonation in the IRMS ion source \(Anicich, 1993\)”.](#)

P16 L8: “CO<sub>2</sub> present” to “presence of CO<sub>2</sub>”

[Section re-worded.](#)

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

[Section re-worded.](#)

P16 L10: “don’t” to “do not”

[Section re-worded.](#)

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

[Section re-worded.](#)

P16 L12: “methane” to “CH<sub>4</sub>” (2 places)

[OK, for all of manuscript.](#)

P16 L14: “quantitative” Does it mean combustion efficiency of 100% or close to 100% without measureable isotopic fractionation? Write explicitly. Meaning of “quantitative” is unclear also for many other places.

[Details provided.](#)

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 CO<sub>2</sub> conversion resulting in high-precision  $\delta^{13}\text{C}$  measurements (...).”

[OK.](#)

P16 L16: “...oxygen isotope composition is altered completely in the conversion process from the original carbonate to the product CO<sub>2</sub>.”

[Section re-worded.](#)

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

[Section re-worded.](#)

P16 L18: “ambiguity” to “uncertainty”;

[OK.](#)

“extracting” to “calculating”;

[OK.](#)

“values” to “value”;

[We leave “values” and use plural in “ion currents” thereafter.](#)

“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 <sup>17</sup>O 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.

[Section re-worded.](#)

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

[“... \$\delta^2\text{H}\$  analyses,...”](#)

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

[“...to produce high conversion yields of ...”](#)

P16 L23: “methane” to “CH<sub>4</sub>”;  
OK for all.

My suggestion: “Major artifact can arise from more variable surface adhesion of H<sub>2</sub>O than CH<sub>4</sub> in the combustion furnace before they are converted to H<sub>2</sub> (and CO/carbon).”

The suggested sentence is misleading. We discuss this in detail as this highlights the importance of the information in Appendix A, which might otherwise be under-appreciated: *i*) The surface adhesion effects of H<sub>2</sub>O are not variable at constant reactor temperatures and constant H<sub>2</sub>O amounts, in fact they are a reproducible function of T, which is why this system works for the calibration of H<sub>2</sub>O against H<sub>2</sub>O. This is demonstrated in Appendix A, Figure A2, where the variability in  $\delta^{2}\text{H-H}_2\text{O}$  is clearly reactor temperature dependent but reproducible for every temperature. Note this temperature effect is insignificant for CH<sub>4</sub>, as CH<sub>4</sub> shows no significant adhesion to surfaces of the analytical system. *ii*) Significant adhesion effects of H<sub>2</sub>O occurs at several places, inside the reactor and on the septum, as discussed in Appendix A. *iii*) This paragraph is not about the combustion reactor but about the pyrolysis occurring in the classy carbon reactor of the TC/EA-IRMS system at temperatures much higher than combustion temperatures and in the absence of oxygen. *iv*) The artefacts range from very strong in case of reactor temperature (~15 %) to significant in the case of septum temperature (~3 %) over the tested temperature ranges, which is why we use the term “possible artefacts”.

We change this sentence in the revised version to “Possible artefacts can arise mainly from the stronger surface activities of H<sub>2</sub>O versus CH<sub>4</sub> prior to the conversion to H<sub>2</sub> (and CO or carbon).”.

P16 L24: “water” to “H<sub>2</sub>O”;  
OK for all places.

I would write: “This can lead to memory effect in the  $\delta^{2}\text{H-H}_2\text{O}$  measurements, then corrections or discarding initial injections are needed (...)"

The suggested sentence is misleading. In fact, the memory effect is induced by H<sub>2</sub>O adhesion to the surfaces of the analytical system (reactor, septum). The artefact due to system memory between injections of the same H<sub>2</sub>O sample decreases with every injection until it is insignificantly small. If we inject a sequence of H<sub>2</sub>O samples, followed by a sequence of CH<sub>4</sub> samples, then the first  $\delta^{2}\text{H-CH}_4$  analysis may in some cases be affected by the memory effect caused by H<sub>2</sub>O adhesion due to desorbing H<sub>2</sub>O. However, that was not consistent.

We change this sentence in the revised version to “H<sub>2</sub>O injections can lead to memory effects, which need to be taken into account in  $\delta^{2}\text{H-H}_2\text{O}$  and subsequent  $\delta^{2}\text{H-CH}_4$  analyses, either by discarding initial injections or by correcting for it (Werner and Brand, 2001)”.

P16 L26: “In addition, we found a minor dependence of...”;  
OK.

“In the appendix” to “In Appendix A”  
OK.

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 CH<sub>4</sub> gases Fossil and Biogenic (Table 3). Both gases...”

**This paragraph is removed from this section in the revised manuscript.**

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

**This paragraph is removed from this section in the revised manuscript.**

P16 L31: “methane” to “CH<sub>4</sub>”; “sampling of”;

**This paragraph is removed from this section in the revised manuscript.**

“consecutive” to “subsequent”?

**This paragraph is removed from this section in the revised manuscript.**

P16 L31–33: “Sperlich et al. (2012) analyzed the CH<sub>4</sub> derived CO<sub>2</sub> for  $\delta^{13}\text{C-CH}_4$  on a dual inlet IRMS and the CH<sub>4</sub> derived H<sub>2</sub>O for  $\delta^2\text{H-CH}_4$  on a TC/EA-IRMS system similar to this study or cavity-ring-down spectroscopy.”

**OK, sentence is moved into method Section 2.9.**

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).”

**Section re-worded.**

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

**Section re-worded.**

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).

**Section re-worded and further developed.**

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.

**See comment above.**

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.

[See comment above.](#)

P17 L11: "uncertainty" to "uncertainties"

[See comment above.](#)

P17 L12: "an indicator"; "create" to "cause"

[See comment above.](#)

P17 L14: I do not understand what the authors argue here. With "CH<sub>4</sub> 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?

[See comment above.](#)

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."

[This paragraph is removed from this section in the revised manuscript, as this is not the case when the uncertainty of LSVEC is considered.](#)

P17 L21: "The number" to "Number"

[?](#)

P17 L22: "could lead to" to "provides";

["...and combining data from multiple laboratories could enable new science and increasingly powerful analysis."](#)

delete "of the combined data sets"

[OK.](#)

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."

["However, merging data from multiple laboratories for analysis is currently hampered by the lack of reference materials that enable the community to produce a unified data set."](#)

P17 L25–26: The paragraph can just follow the previous one without line break;

[OK.](#)

My suggestion: "To deal with this problem, we prepared 12 pure CH<sub>4</sub> gases (the secondary CH<sub>4</sub> 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 CH<sub>4</sub> gases then were diluted to produce 8 synthetic CH<sub>4</sub>-in-air standards in 5-L glass flasks."

["To overcome this problem and to improve compatibility between laboratories, we produced synthetic CH<sub>4</sub>-in-air standards \(JRAS-M16\)."](#)

P17 L26–28: I do not understand what the authors argue here. The authors say the synthetic CH<sub>4</sub>-in-air standards were “tested for their use”, but where in the manuscript did they evaluate usability of the gases? Section 3.3 does not seem to describe this. “separately” to “indendently”.

[Section reformulated in revised manuscript.](#)

P17 L29: “synthetic atmospheric reference gases” to “synthetic CH<sub>4</sub>-in-air standards”; “isotopic composition of CH<sub>4</sub>” to “ $\delta^{13}\text{C-CH}_4$  and  $\delta^2\text{H-CH}_4$ ”

[OK.](#)

P17 L30: My suggestion: “These synthetic CH<sub>4</sub>-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 CH<sub>4</sub> cycle.”

[Section reformulated in revised manuscript.](#)

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

[“Experiments to enhance the performance of the analytical system for the calibration of  \$\delta^2\text{H-CH}\_4\$  with H<sub>2</sub>O”.](#)

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

[We disagree. The effect is still there, but it is overwritten or overcome by repeated injections.](#)

P18 L13: “However” to “Moreover”

[OK.](#)

P18 L14: delete “furthermore”;

[OK.](#)

“scales with isotopic difference” to “depends on difference of isotope ratio”

[Matter of taste, we prefer the original version.](#)

P18 L15: delete “only”

[OK.](#)

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

[Section reformulated in revised manuscript.](#)

P18 L17: delete “for” after “corrected”

[We think “corrected for” is correct.](#)

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.

While we accept the argument for the Kr example (reviewer's comment to previous P17 L5), we disagree in this case as both cases are not comparable. The Kr effect was possible due to the lack of *i*) knowledge and *ii*) suitable reference materials. In the case discussed here, knowledge, suitable reference materials and experimental methods are available to guarantee sufficient control on system memory during  $\delta^2\text{H-H}_2\text{O}$  analysis.

P18 L21: “We made 106 injections of an identical  $\text{H}_2\text{O}$  samples...”  
“We injected 106 identical  $\text{H}_2\text{O}$  samples”.

P18 L22: “Septum” to “septum”;  
OK.

I would delete “however, there seems to...”  
OK.

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.”

“A systematic increase of  $\delta^2\text{H-H}_2\text{O}$  with septum temperature is apparent above 90°C until  $\delta^2\text{H-H}_2\text{O}$  values plateau at septum temperatures 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.

We removed the sentence as we do not have values for 130°C exact.

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

Matter of taste, we prefer the original version.

P18 L26: “at the lower temperature range” to “below 90°C”;  
OK.

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?

OK.

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

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”.

OK.

P19 L4: What is “the offset”?

Sentence re-worded.

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

Reformulated: “The error bars show  $1\sigma$  standard deviations and the grey-dashed lines indicate the typical precision limit of 1 ‰ for  $\delta^2\text{H}-\text{H}_2\text{O}$  analysis (Gehre et al., 2004) around the  $\delta^2\text{H}-\text{H}_2\text{O}$  value of the polynomial fit for the septum temperature of 130°C (set point during calibration experiments). The grey dashed lines show that our  $\delta^2\text{H}-\text{H}_2\text{O}$  analyses remain within a typical precision level as long as the septum temperature is controlled to  $\sim 130 \pm 10^\circ\text{C}$ .”

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

OK.

“utmost” to “particular”

Matter of taste, we prefer the original version.

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

? “The reactor temperature is...”

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

OK.

P19 L16: “150 K” to “150°C”

OK.

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

We included the experiments of Appendix B into the discussion part of the main text.