

General Comments:

First, I want to congratulate Sperlich and co-workers for their extremely valuable and important work, which will hopefully, more than three decades after the first publications of isotope measurements in atmospheric methane, solve our problem of lack of reference standards for these analyses. Sperlich et al. present a sound way for linking carbon and hydrogen isotope ratios in pure CH₄ to the internationally accepted IAEA carbonate and water reference materials. After dilution of these calibrated pure CH₄ gases with CH₄-free synthetic air they produce CH₄-in-air mixtures of ambient concentrations that can be used in the future as calibration standards, linking atmospheric (and source) methane isotope analyses from globally distributed labs to a common calibration scale.

As was already pointed out by Referee # 1, this fundamental work will become one of our basic references to describe the development of our future methane isotope calibration scale. As such, however, the descriptions of procedures in the current version of the manuscript, unfortunately, do not fully meet the requirements for clarity and completeness. Referee # 1 has already prepared a long list of comments and made very good suggestions for improvements of the manuscript, which I fully support. In my list of comments below, I thus only want to re-emphasize a number of points, which I feel most important to be tackled in a revised manuscript.

Specific Comments:

1. Introducing the various standard materials, their production (e.g. also by spiking with deuterated CH₄), hierarchy and their calibration against IAEA reference materials (rm), or against other CH₄ gases or other CH₄-in-air gas mixtures is rather confusing. This does not only concern Figure 1 and Tables 1 and 2, but also Table 3, where the calibration results are given. I like very much the revised Figure 1 suggested by Referee #1. Please also be VERY clear with your nomenclature, e.g. distinguishing between “calibrations” (i.e. against reference materials) and what, from my point of view should better be named “comparison” with the earlier MPI-Jena standard gases Carina. In fact, it is not really clear to me how the H₂ scale in Brass and Röckmann (2010), which forms the basis of the earlier MPI-Jena scale has been established. In their paper Brass and Röckmann refer to a paper by Bergamaschi et al. (1994) who obtained their calibration from colleagues at BGR, Hannover. The observed $\delta^2\text{H}$ difference of 4‰ between the IMAU/earlier MPI-BGC scale and the recent calibration may perhaps not be surprising. What does the remark on page 12 line 24-25 mean in this context? More information about the origin of the IMAU/earlier MPI-BGC scale is required to judge on the comparison results listed in Table 4.
I think, the sentence in the conclusion P17, L 27 is too strong as the earlier MPI-BGC scale is only propagated from some yet unexplained origin.
2. Concerning the experimental set up for the calibrations against carbonates and water, I find the descriptions confusing and much too brief. A figure that displays the complete setups (for CH₄ against carbonates and for CH₄ against water) would be very helpful. Figures 2 and 4 could then be integrated there.
3. Discussion on accuracy of the calibrations: Although the authors have explained in detail how they tried to follow, as much as possible, the principle of identical treatment (PIT) and to avoid possible pitfalls when calibrating CH₄ against carbonate and H₂O reference materials, they cannot be sure that indeed no systematic biases have occurred. The most convincing argument for accuracy of the new standards to me is the good agreement with the earlier work by Sperlich et al. (2012) who used a (slightly) different procedure than in the present work. The discussion of the uncertainty in this respect is not clear enough. It seems rather to come as a mixture of long-term precision, agreement with the IMAU scale (see my reservations above) and finally arguing with “the combined uncertainty”. I would like to see

here a more elaborated discussion and clear separation of the different indicators for accuracy. This could hopefully help to pin down biases in the future.

4. I also agree with Referee #1 that a description (or at least a reference to a publication) of the iSAAC measurement system is required.

Minor comments:

The Appendix is named Appendix 1 in the text but A and B when they show up

Sect. 3.2: that the high d2H values have been produced by spiking should go into section 2.1

In the discussion section it may be helpful to explain why $m/z = 15$ is used to detect unconverted CH₄ in the sample.