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 δ^{13}C-CH_4 and δ^{2}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 δ^{13}C-CH_4 and δ^{2}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_4-in-air standards and to the pure CH_4 gases as primary and secondary CH_4 gases. "Working standard" is abbreviated WS, "certified reference material" CRM, "reference material" RM and "matrix reference material" (e.g. CH_4 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_4➔secondary CH_4). 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 2 (Ingeborg Levin)

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 CH_4 to the internationally accepted IAEA carbonate and water reference materials. After dilution of these calibrated pure CH_4 gases with CH_4-free synthetic air they produce CH_4-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 CH_4), hierarchy and their calibration against IAEA reference materials (rm), or against other CH_4 gases or other CH_4-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.
OK, a similar figure has been added to the revised manuscript. We hope this clarifies the hierarchies and relations between applied gases and reference materials.

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.
OK. We define our use of calibration and measurement at the beginning of the method section.

In fact, it is not really clear to me how the H_2 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.
We dedicate a new section on the description of the scale history at IMAU.

The observed δ^2H difference of 4‰ between the IMAU/earlier MPI-BGC scale and the recent calibration may perhaps not be surprising.
That is true, it is not surprising. It is rather surprisingly good considering the previously achievable measurement precisions and scale propagations. We think this is better discussed in the revised manuscript.

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.
OK, detailed information are provided in revised manuscript.

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.
OK, section is re-formulated and considers the concerns on propagated scale anchor.

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 CH_4 against carbonates and for CH_4 against water) would be very helpful. Figures 2 and 4 could then be integrated there.
New figures are included in the manuscript that show the reactors of the TC/EA-IRMS and the EA-IRMS system. Both figures contain the 10-port valve configuration to display the CH_4 injection into each of the systems. All other reference materials are introduced via autosampler in both systems. We think the descriptions in the revised manuscript are more clear and detailed.

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 CH_4 against carbonate and H_2O 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.
OK. The revised manuscript has a strong focus on the comparison with the earlier work of Sperlich et al., (2012). In fact, we present results from new experiments to compare the methods from CIC and MPI-BGC in more detail. In order to do so, we revised the data evaluation and the uncertainty calculation from Sperlich et al., (2012). We agree this part has fallen short in the previous version of the manuscript and think it is much clearer in the revised version. While we address the differences in the comparison, our explanation for the cause of the differences is limited. We discuss the potential for measurement artefacts, such as incomplete conversion or scale compression effects. However, we discuss why we think that we have good control on these processes and that we can therefore not identify the cause of the inter-laboratory differences.

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.
OK, included. As mentioned above, the respective paper has just been published (Brand et al., 2016).

Minor comments:

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

Sect. 3.2: that the high δ^2H values have been produced by spiking should go into section 2.1
OK.

In the discussion section it may be helpful to explain why *m/z* = 15 is used to detect unconverted CH_4 in the sample.
OK, explanation included in revised manuscript.