Stable isotope measurements of greenhouse gases CO_2 and CH_4 make a powerful tool used to understand processes involved in the global carbon cycle. In order to get meaningful interpretation of stable isotope data in greenhouse gases, the data produced in different labs and in different years should be compatible within certain limits (WMO, 2014), these are called as compatibility goals (Table 1). In the last years compatibility of air-CO_2 stable isotope data is thought to be improved by introducing calibration “JRAS air” mixtures (WMO, 2012).

The compatibility goals for CH_4 are still a challenge to be achieved; it can be realized by using optimised calibration schemes, and to be based on appropriate reference materials with low uncertainty. All in all the compatibility goals can be considered in the first instance as the uncertainty required for “fit-for purpose” reference materials. In turn, such reference materials have to be compatible with a sample, that is why the community needs several reference CH_4-in-air mixtures.

Table 1. Compatibility goals for atmospheric CH4 (after WMO, 2014).

Component	**Compatibility goal (for background air)**	**Extended compatibility goal (for polluted air)**
$\delta^{13}C$-CH_4	± 0.02‰	± 0.2‰
$\delta^{2}H$-CH_4	± 1‰	± 5‰

Thereafter, work on the calibration of pure CH_4 gases aimed to produce reference CH_4-mixures cannot be published without thoughtful considerations of the uncertainty estimation and without clear presentation of the uncertainty budget. In particular this is expected for the work presented by the WMO-GAW Central Calibration Lab for stable isotopes in greenhouse gases (currently MPI-BGC, Jena, DE). In this respect the manuscript demonstrates serious problems such as unclear presentation of the calibrating approach in general, unclear uncertainty budget, as well as potentially missing/neglecting some uncertainty components. The major shortcuts are as following:

1. In order to build a skeleton of the uncertainty propagation, one has to consider a traceability chain for all measurements. The traceability chain has to be tracked to the highest Ref.Materials (RMs) in use. In case of $\delta^{13}C$ these are NBS19 & LSVEC (these have to be considered with their uncertainties) and include all measurement steps. Each next measurement step (including measurements on RMs) introduces an analytical uncertainty, thus increasing the total uncertainty.
2. The uncertainty propagation should be based on the traceability chain and also include all potential effects due to TC/EA, gas dilution etc. Besides, I would suggest to present the uncertainty budget, namely to describe a contribution of each uncertainty component starting from the uncertainty assigned to RMs carbonates (NBS19 and LSVEC), then the uncertainty of carbonate measurements, the uncertainty of 2-poit calibration as based on the carbonates, then analytical uncertainty of “master”-CH_4 (this is used for calibration “calibration”CH_4) etc. Such uncertainty budget will clearly demonstrate where further improvements are essential.
3. The “master” CH_4 (and its replacement when the first “master” was lost) was calibrated vs the IAEA Ref.Materials by applying the 2-point calibration approach. Next, several “calibration” CH_4 were calibrated vs the “master” CH_4. It is unclear how the 2-point calibration was applied in the case of measuring several “calibration” CH_4 gases? In fact

calibration vs. the “master” CH_4 looks like 1-point, thus violating the 2-point calibration approach (Coplen et al., 2006) designed to address various effects during sample preparation and measurements. I stress – this is in particular critical for $\delta^{13}C$ values being down to -69.9 ‰ (Tab 3 in the manuscript), far below -40 ‰ of the “master”CH_4 and also outside the LSVEC value of -46.6 ‰.

4. Given that “calibration” CH_4 gases were characterised against the “master” CH_4, it is unclear why the $\delta^{13}C$-uncertaitnty of 0.06 ‰ for Martha-1 (“calibration”-CH_4) is smaller than the uncertainty of 0.07 ‰ obtained for the “master” CH_4. The uncertainty of each next material cannot be smaller than the uncertainty of material(s) used for its calibration (in this case uncertainty of “master” CH_4). This example implies something to be wrong in the uncertainty evaluation scheme in general. For the same reasons the $\delta^{13}C$ uncertainty of ± 0.08 ‰ given for the “calibration” CH_4 Mike-1 looks like optimistically too low.
5. The authors should also explain the uncertainty values for “Biogenic” and “Fossil” CH_4, namely the values of ± 0.04‰, as given with the reference to (Sperlich et al., 2012). Why these are lower than uncertainties obtained by the work presented in this manuscript? In fact Sperlich et al. (2012) gave no detailed explanation on the uncertainty propagation. Given that the paper by Sperlich et al. (2012) is written by the same authors as the present manuscript, this is a must-requirement.
6. When focusing high accuracy values, the authors need to consider the effect ^{17}O correction for the entire $\delta^{13}C$-calibration scheme, namely when calibration started from carbonates is applied to CH4 gases. Is there any potential bias?
7. Last but not least, the authors wrongly cite the $\delta^{13}C$-uncertainty of LSVEC. The message sent in Dec-2016 by the IAEA to LSVEC customers suggests the $\delta^{13}C$-uncertainty of LSVEC at ±0.15 ‰; this value is also used by A.Schimmelmann et al. 2016 (see http://pubs.acs.org/doi/abs/10.1021/acs.analchem.5b04392). The present interpretation of the message distributed by the IAEA is misleading.

All in all I find the uncertainty evaluation presented in the manuscript as unclear, confusing and partly misleading. The uncertainty evaluation for δ^2H may suffer for similar reasons. Given the problem with LSVEC, the $\delta^{13}C$ uncertainty presently achieved appears not fulfilling the requirements.

Sergey ASSONOV (reference material specialist for stable isotopes)
IAEA Environment Laboratories, IAEA