Development and evaluation of a suite of isotope reference gases for methane in air

Sperlich, Peter; Uitslag, Nelly A. M.; Richter, Juergen M.; Rothe, Michael; Geilmann, Heike; van der Veen, Carina; Rockmann, Thomas; Blunier, Thomas; Brand, Willi A.

Published in:
Atmospheric Measurement Techniques

DOI:
10.5194/amt-9-3717-2016

Publication date:
2016

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Development and evaluation of a suite of isotope reference gases for methane in air

Peter Sperlich1,2, Nelly A. M. Uitslag1,3, Jürgen M. Richter1, Michael Rothe1, Heike Geilmann1, Carina van der Veen4, Thomas Röckmann4, Thomas Blunier5, and Willi A. Brand4

1Max Planck Institute for Biogeochemistry (MPI-BGC), Jena, Germany
2National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand
3Centre for Isotope Research (CIO), Energy and Sustainability Research Institute Groningen (ESRIG), University of Groningen, Groningen, the Netherlands
4Institute of Marine and Atmospheric Science in Utrecht (IMAU), Utrecht, the Netherlands
5Centre for Ice and Climate (CIC), Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

Correspondence to: Peter Sperlich (peter.sperlich@niwa.co.nz)
Received: 18 January 2016 – Published in Atmos. Meas. Tech. Discuss.: 20 January 2016
Revised: 10 June 2016 – Accepted: 8 July 2016 – Published: 12 August 2016

Abstract. Measurements from multiple laboratories have to be related to unifying and traceable reference material in order to be comparable. However, such fundamental reference materials are not available for isotope ratios in atmospheric methane, which led to misinterpretations of combined data sets in the past. We developed a method to produce a suite of synthetic CH4-in-air standard gases that can be used to unify methane isotope ratio measurements of laboratories in the atmospheric monitoring community. Therefore, we calibrated a suite of pure methane gases of different methanogenic origin against international referencing materials that define the VSMOW (Vienna Standard Mean Ocean Water) and VPDB (Vienna Pee Dee Belemnite) isotope scales. The isotope ratios of our pure methane gases range between −320 and +40 ‰ for δ2H–CH4 and between −70 and −40 ‰ for δ13C–CH4, enveloping the isotope ratios of tropospheric methane (about −85 and −47 ‰ for δ2H–CH4 and δ13C–CH4 respectively). Estimated uncertainties, including the full traceability chain, are < 1.5 ‰ and < 0.2 ‰ for δ2H and δ13C calibrations respectively. Aliquots of the calibrated pure methane gases have been diluted with methane-free air to atmospheric methane levels and filled into 5 L glass flasks. The synthetic CH4-in-air standards comprise atmospheric oxygen/nitrogen ratios as well as argon, krypton and nitrous oxide mole fractions to prevent gas-specific measurement artefacts. The resulting synthetic CH4-in-air standards are referred to as JRAS-M16 (Jena Reference Air Set – Methane 2016) and will be available to the atmospheric monitoring community. JRAS-M16 may be used as unifying isotope scale anchor for isotope ratio measurements in atmospheric methane, so that data sets can be merged into a consistent global data frame.

1 Introduction

Isotope ratios of CH4 in the present and the past atmosphere (e.g. from ice cores) are a powerful tool to study the biogeochemical processes that cause the variation of CH4 in the atmosphere (Stevens and Rust, 1982; Quay et al., 1991, 1999; Lowe et al., 1994; Sapart et al., 2012; Möller et al., 2013; Sperlich et al., 2015; Schaefer et al., 2016). Recently, two conflicting publications highlighted (i) the interpretative power when data sets from multiple laboratories are combined for spatiotemporal analysis of CH4 isotope ratios (Kai et al., 2011) and (ii) the pitfalls when differences due to laboratory offsets are misinterpreted as spatial variability of CH4 sources (Levin et al., 2012). Levin et al. (2012) identified calibration offsets between three laboratories by comparing their long-term observations in Antarctic background air, where the δ13C of CH4 is assumed to be free of spatial gradients. However, this technique is a temporary work-around that excludes the use of data sets from laboratories without a history of observations in Antarctica or a traceable link to
Antarctic observations. This dilemma could be solved if suitable reference materials (RMs) were available to all laboratories that measure isotope ratios of atmospheric CH₄.

Certified reference materials (CRMs) are provided by the IAEA, NIST and others for many analytes. The lack of CRMs for CH₄ isotope ratios has long been recognised in the literature, ranging from pioneering papers (e.g. Craig, 1953; Schiegl and Vogel, 1970) to recent publications on analytical systems to measure isotope ratios in atmospheric CH₄ (e.g. Sapart et al., 2011; Sperlich et al., 2013; Bock et al., 2014; Tokida et al., 2014; Eyer et al., 2015) as well as papers that present and interpret such data (e.g. Levin et al., 2012; Sapart et al., 2013; Schaefer et al., 2016). In the absence of CRMs for isotope ratios of CH₄, many laboratories have developed methods to calibrate purified CH₄ against CRMs that were available as a “second-best solution”, thereby accepting the shortcoming that those CRMs comprised of different physicochemical properties and are therefore not ideal (IAEA, 2003). For example, δ¹³C–CH₄ calibrations were made against NBS 20 (limestone) and NBS 21 (graphite) by Stevens and Rust (1982), against NBS 16 (CO₂) and NBS 20 (limestone) by Quay et al. (1991), against IAEA-CO-9 (Barium carbonate) by Lowe et al. (1994), against NBS 19 (limestone) by Quay et al. (1999) and against RM 8563 (CO₂) by Sperlich et al. (2012). Dumke et al. (1989) calibrated against the natural gas mixtures NGS 1, NGS 2 and NGS 3, which were not of the highest purity level with 81, 53 and 99 % CH₄ respectively (e.g. IAEA, 2003; Brand et al., 2014). It is furthermore important to understand the variation of uncertainties of the applied CRMs, ranging from assigned values of 0.00 ‰ (NBS 19, the only primary measurement standard for VPDB) up to 0.56 ‰ (NGS 2) (Brand et al., 2014). The situation becomes even more complicated because the δ¹³C values of some of the applied CRMs were revised and changed by as much as 0.4 ‰ over time (e.g. NBS 21; Brand et al., 2014). As a consequence, this would require the adjustment of dependent δ¹³C–CH₄ data. The use of different calibration methods, CRMs and the change of their assigned δ¹³C values have undoubtedly contributed to calibration offsets between laboratories. This fact highlights the importance that applied CRMs and their δ¹³C values are reported in the metadata of the measurement results and that their uncertainty is included in the uncertainty budget of the measurements. Fortunately the situation is more homogenous for δ²H–CH₄ calibrations, which were only made against CRM waters, such as VSMOW2, SLAP2 or their precursors (e.g. Schiegl and Vogel, 1970; Dumke et al., 1989; Quay et al., 1999; Sperlich et al., 2012). Brand et al. (2014) provide a comprehensive overview on the variation of δ²H–H₂O values and associated uncertainties. Another common method for laboratories to anchor CH₄ measurements to the VPDB or VSMOW isotope scales is to get their working standard (WS) calibrated by an external laboratory (e.g. Behrens et al., 2008; Brass and Röckmann, 2010; Bock et al., 2014; Schmitt et al., 2014; Rella et al., 2015; Brand et al., 2016). It is important to keep in mind that propagating isotope scales between laboratories also requires inclusion and propagation of the uncertainty of the respective isotope scale anchor.

In summary, the absence of unique CRMs for δ²H–CH₄ and δ¹³C–CH₄ led to a diversity of calibration trajectories. Significant calibration offsets between laboratories on the order of 0.05–0.09 ‰ for δ¹³C–CH₄ were identified through co-located measurements by Levin et al. (2012) and Schaefer et al. (2016), while Bock et al. (2014) reported laboratory offsets of up to 15 ‰ for δ²H–CH₄. Even though inter-laboratory differences can be established experimentally, e.g. by co-located measurements or regular round robins, such comparisons are not intended to re-define local scale anchors to the VPDB and VSMOW isotope scales (WMO, 2014) and can therefore not replace a unifying scale anchor.

Until recently, a comparable problem existed for observations of isotope ratios in atmospheric CO₂. Ghosh et al. (2005) established a method to produce synthetic CO₂-in-air standards, comprising of isotopically calibrated CO₂ and CO₂-free air. The concept of these CO₂-in-air standards is to provide a matrix reference material (m-RM), which is defined as RM that is mixed with matrix material to match the composition of the samples (IAEA, 2003). Since 2005, the ISOLAB of the Max Planck Institute for Biogeochemistry (MPI-BGC) in Jena, Germany, distributes a suite of m-RMs, known as JRAS (Jena Reference Air Set), which is accepted as an isotope scale anchor by the community (WMO, 2012). Calibrating against the JRAS reduces laboratory offsets and has proven a successful method to reach and maintain the compatibility goal for isotope ratios in atmospheric CO₂ (Wendeberg et al., 2013).

This paper describes an analogue method to produce synthetic CH₄-in-air standards for δ²H–CH₄ and δ¹³C–CH₄, which we refer to as JRAS-M16 (short for JRAS-Methane 2016). We present new methods to calibrate a suite of isotopically different CH₄ gases, which span over a large isotopic range. We calibrate two CH₄ gases for δ²H and δ¹³C and compare our results to independent calibrations made at a partnering laboratory to demonstrate the comparability of our new methods, thereby fulfilling the requirement to use two independent analytical methods during the development of quality control materials (QCMs) when CRMs are not available (IAEA, 2003). We produce synthetic CH₄-in-air standards by diluting aliquots of calibrated CH₄ with CH₄-free synthetic air and include the full traceability chain in the uncertainty budget. Calibrated δ²H–CH₄ and δ¹³C–CH₄ values in our synthetic CH₄-in-air standards bracket tropospheric values and enable two-point calibrations to account for scale compression effects (Coplen et al., 2006a). Our synthetic CH₄-in-air standards can be tested by other laboratories in the community; alternatively, compressed air cylinders from other laboratories can be calibrated at MPI-BGC. Our long-term strategy is to establish JRAS-M16 as m-RM for δ²H–CH₄ and δ¹³C–CH₄ in the future. We hope that our efforts help the community to reach the scale anchor compat-

Atmos. Meas. Tech., 9, 3717–3737, 2016 www.atmos-meas-tech.net/9/3717/2016/
ability goals of 1 and 0.02 ‰ for $\delta^2$H–CH$_4$ and $\delta^{13}$C–CH$_4$ respectively (WMO, 2014).

2 Materials and methods

Throughout this paper, we use the terminology of “calibration” and “measurement” with different intentions. We use calibration when samples are repeatedly compared against measurement standards of the highest possible hierarchy level (possible hierarchy levels include CRMs and WSs) in order to determine the isotopic composition of the analyte under consideration of the full traceability chain. In contrast, we use the measurement term when the analysis is not necessarily based on measurement standards of highest possible hierarchy level, when the achievable uncertainty of the analysis is not of primary importance or when the uncertainty does not necessarily include the full traceability chain. For example, we use the measurement term for the experiments to establish the dependence of isotope ratios in the analyte on reactor temperatures of the analytical system.

The aim of our method is to calibrate and prepare synthetic CH$_4$-in-air standards, as outlined in the flow diagram of Fig. 1. Therefore, we calibrate two pure CH$_4$ gases for their $\delta^2$H–CH$_4$ and $\delta^{13}$C–CH$_4$ isotope ratios against CRMs and WSs, where the latter are of comparable chemical composition to the former. We refer to these two CH$_4$ gases as primary CH$_4$ gases. The primary CH$_4$ gases are then used to calibrate a suite of pure CH$_4$ gases, which we refer to as secondary CH$_4$ gases. The analytical methods we developed for $\delta^2$H–CH$_4$ and $\delta^{13}$C–CH$_4$ calibrations are based on well-established IRMS methods, thereby complying with the requirements to use established analytical systems for the production of QCMs when CRMs are not available (IAEA, 2003). Once calibrated, aliquots of both primary and secondary CH$_4$ gases are diluted with CH$_4$-free air to atmospheric CH$_4$ mole fractions. We analyse the resulting synthetic CH$_4$-in-air standards on a new analytical system that is designed to analyse atmospheric samples, thereby complying with the principle of identical treatment (PIT; Werner and Brand, 2001) during the analysis of the synthetic CH$_4$-in-air standards. This enables us to determine the calibration difference between JRAS-M16 and the hitherto adopted method to reference $\delta^2$H–CH$_4$ and $\delta^{13}$C–CH$_4$ in atmospheric samples to the VSMOW and VPDB scales respectively. This difference represents the laboratory specific correction that has to be applied to anchor all measurements from MPI-BGC to the new JRAS-M16 scale.

2.1 Gases, reference materials and hierarchy levels of calibrations

Our study is based on a suite of CH$_4$ gases that differ in their methanogenic origin and therefore in their isotopic composition. We identify our CH$_4$ gases by names as shown in Table 1. “Biogenic” and “Fossil” have been calibrated at the Centre for Ice and Climate (CIC), which is a department of the Niels Bohr Institute at the University in Copenhagen, Denmark (Sperlich et al., 2012). These gases allow testing and evaluating the performance of the analytical systems at MPI-BGC with independent methods, which is a required control mechanism for the development of QCMs when CRMs are not available (IAEA, 2003). Six other CH$_4$ gases were purchased from suppliers of commercial gases or laboratory equipment (Air-Liquide, Westfalen AG, Linde, Messer, Campro Scientific) and were used as purchased or as mixtures thereof. The purity level of all our CH$_4$ gases is
≥ 99.995%. Our goals were to produce (i) a suite of CH$_4$ gases that encompasses the isotopic composition of tropospheric CH$_4$, and (ii) CH$_4$ gases that closely match the isotopic composition of tropospheric CH$_4$. For $\delta^2$H–CH$_4$ this was achieved by spiking fossil CH$_4$ gases with pure CH$_3$D to yield “Martha-1”, “Martha-2” and “Mike-1”. Mike-1 was then mixed with a fossil CH$_4$ gas to produce “Mike-2” while Martha-1 was spiked with pure CH$_3$D to produce “Martha-2”. Martha-1 and Mike-1 were thereby transitional CH$_4$ mixtures.

We calibrated “Megan” and “Merlin” for $\delta^2$H against VSMOW/SLAP and against other pure CH$_4$ gases

2.2 Referencing pure CH$_4$ for $\delta^2$H against VSMOW/SLAP and against other pure CH$_4$ gases

We use a high-temperature conversion elemental analyser (TC/EA) coupled to an isotope ratio mass spectrometer (IRMS; Delta Plus XL, Thermo Finnigan, Bremen, Germany) via an open split (ConFlo III, Thermo Finnigan, Bremen, Germany). The system at MPI-BGC is operated for $\delta^2$H–H$_2$O and $\delta^{18}$O–H$_2$O analysis with high precision and negligible systematic errors since more than a decade (Gehre et al., 2004; Brand et al., 2009a) and is depicted in Fig. 2. Because TC/EA–IRMS systems are also used for $\delta^2$H analysis in hydrocarbons (e.g. Hilkert et al., 1999; Schimmelmann et al., 2016), this method is particularly suitable to calibrate $\delta^2$H–CH$_4$ against reference H$_2$O (CRM and WS, Table 2).

Therefore, we inject CH$_4$ and H$_2$O samples through an externally heated septum (kept at 130°C) into the glassy carbon reactor of the TC/EA–IRMS (kept at 1450°C), where both species are converted to H$_2$ (+ carbon or CO). A helium carrier gas transports the sample gases from the high-temperature reactor through a gas chromatographic (GC) column (1/4 in. OD, 60 cm length, 5-Å zeolite, 75°C) and into the open split. CH$_4$ injections are made with a two position 10-port valve (VICI, USA), which is configured as depicted in Fig. 2. A helium stream of 15 mL min$^{-1}$ carries CH$_4$ sam-

Figure 2. Configuration of manual the two-position 10-port valve with two 1 mL sample loops shown in grey dashed box and TC/EA–IRMS system for $\delta^2$H–CH$_4$ calibration. The TC/EA–IRMS reactor (displayed as in Gehre et al., 2004) is fed either by the sample line from the 10-port valve or by the syringe via autosampler (not shown). The size of components is chosen to increase clarity.
Table 1. Gases used for this study. Note that Mike-1 and Martha-1 were transitional CH$_4$ mixtures and do not exist anymore.

<table>
<thead>
<tr>
<th>Cylinder name</th>
<th>Volume (L)</th>
<th>Pressure (bar)</th>
<th>Function in study</th>
<th>CH$_4$ source</th>
<th>Gas supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Megan</td>
<td>10</td>
<td></td>
<td>first primary CH$_4$ (lost)</td>
<td>fossil CH$_4$</td>
<td>Air Liquide, Germany</td>
</tr>
<tr>
<td>Merlin</td>
<td>10</td>
<td>190</td>
<td>second primary CH$_4$ (replacement of Megan)</td>
<td>fossil CH$_4$</td>
<td>Air Liquide, Germany</td>
</tr>
<tr>
<td>Mike-1</td>
<td>–</td>
<td>–</td>
<td>secondary CH$_4$</td>
<td>MPI mixture</td>
<td>MPI-BGC</td>
</tr>
<tr>
<td>Mike-2</td>
<td>5</td>
<td>45</td>
<td>secondary CH$_4$</td>
<td>MPI mixture</td>
<td>MPI-BGC</td>
</tr>
<tr>
<td>Merkur</td>
<td>2</td>
<td>100</td>
<td>secondary CH$_4$</td>
<td>fossil CH$_4$</td>
<td>Messer Griesheim, Germany</td>
</tr>
<tr>
<td>Merida</td>
<td>10</td>
<td>175</td>
<td>secondary CH$_4$</td>
<td>unknown</td>
<td>Westfalen AG, Germany</td>
</tr>
<tr>
<td>Martha-1</td>
<td>–</td>
<td>–</td>
<td>secondary CH$_4$</td>
<td>MPI mixture</td>
<td>MPI-BGC</td>
</tr>
<tr>
<td>Martha-2</td>
<td>10</td>
<td>165</td>
<td>secondary CH$_4$</td>
<td>unknown</td>
<td>Messer Griesheim, Germany</td>
</tr>
<tr>
<td>Minion</td>
<td>3</td>
<td>150</td>
<td>secondary CH$_4$</td>
<td>unknown</td>
<td>Westfalen AG, Germany</td>
</tr>
<tr>
<td>Melly</td>
<td>50</td>
<td>193</td>
<td>secondary CH$_4$</td>
<td>unknown</td>
<td>Westfalen AG, Germany</td>
</tr>
<tr>
<td>δ$_2$H-spike gas</td>
<td>0.4</td>
<td>2.5</td>
<td>CH$_3$D spiking gas</td>
<td>Campro Scientific, Germany</td>
<td></td>
</tr>
<tr>
<td>Fossil</td>
<td>30</td>
<td>2</td>
<td>secondary CH$_4$ and comparison</td>
<td>fossil CH$_4$</td>
<td>Air Liquide, Denmark</td>
</tr>
<tr>
<td>Biogenic</td>
<td>30</td>
<td>2</td>
<td>secondary CH$_4$ and comparison</td>
<td>biogas plant</td>
<td>Biogas Plant, Germany</td>
</tr>
<tr>
<td>Synthetic air</td>
<td>50</td>
<td>200</td>
<td>Synthetic air matrix</td>
<td>Linde, Germany</td>
<td></td>
</tr>
<tr>
<td>Krypton</td>
<td>2</td>
<td>200</td>
<td>synthetic air matrix</td>
<td>Westfalen AG, Germany</td>
<td></td>
</tr>
<tr>
<td>Carina-1</td>
<td>50</td>
<td>200</td>
<td>working standard and scale comparison</td>
<td>Jena air</td>
<td>MPI-BGC</td>
</tr>
<tr>
<td>Carina-2</td>
<td>50</td>
<td>200</td>
<td>working standard and scale comparison</td>
<td>Jena air</td>
<td>MPI-BGC</td>
</tr>
</tbody>
</table>

Table 2. Measurement standards used in this study. “CRM” and “WS” identify certified reference material and in-house working standards respectively. The uncertainties of the δ$^2$H and δ$^{13}$C data from MPI-BGC correspond to the 95 % confidence limit of the error of the mean. We include the uncertainty estimate that the IAEA recently suggested for LSVEC. Publications and additional comments related to the standards are listed in the last column.

<table>
<thead>
<tr>
<th>Name</th>
<th>Material</th>
<th>CRM/WS</th>
<th>δ$^2$H [%e]</th>
<th>δ$^{13}$C [%e]</th>
<th>Source</th>
<th>Reference/comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>VSMOW2</td>
<td>H$_2$O</td>
<td>CRM</td>
<td>0 ± 0.3</td>
<td>–</td>
<td>IAEA</td>
<td>Gröning et al. (2007), Brand et al. (2014)</td>
</tr>
<tr>
<td>SLAP2</td>
<td>H$_2$O</td>
<td>CRM</td>
<td>–427.5 ± 0.3</td>
<td>–</td>
<td>IAEA</td>
<td>Gröning et al. (2007), Brand et al. (2014)</td>
</tr>
<tr>
<td>GISP</td>
<td>H$_2$O</td>
<td>CRM</td>
<td>–189.7 ± 0.9</td>
<td>–</td>
<td>IAEA</td>
<td>Brand et al. (2014)</td>
</tr>
<tr>
<td>NBS 19</td>
<td>CaCO$_3$</td>
<td>CRM</td>
<td>+1.95 ± 0.00</td>
<td>+50.4 ± 0.15</td>
<td>IAEA</td>
<td>Brand et al. (2014), exhausted</td>
</tr>
<tr>
<td>LSVEC</td>
<td>Li$_2$CO$_3$</td>
<td>CRM</td>
<td>–</td>
<td>–46.6 ± 0.15</td>
<td>IAEA</td>
<td>Coplen et al. (2006b), Qi et al. (2016), Schimmelmann et al. (2016)</td>
</tr>
<tr>
<td>CO-9</td>
<td>BaCO$_3$</td>
<td>CRM</td>
<td>–</td>
<td>–47.32 ± 0.06</td>
<td>IAEA</td>
<td>Coplen et al. (2006b)</td>
</tr>
<tr>
<td>RM 8563</td>
<td>CO$_2$</td>
<td>CRM</td>
<td>–</td>
<td>–41.59 ± 0.06</td>
<td>IAEA</td>
<td>Coplen et al. (2006b), exhausted</td>
</tr>
<tr>
<td>WWW-J1</td>
<td>H$_2$O</td>
<td>WS</td>
<td>–67.0 ± 0.4</td>
<td>–</td>
<td>MPI-BGC</td>
<td>–</td>
</tr>
<tr>
<td>BGP-J1</td>
<td>H$_2$O</td>
<td>WS</td>
<td>–187.1 ± 0.6</td>
<td>–</td>
<td>MPI-BGC</td>
<td>–</td>
</tr>
<tr>
<td>MAR-J1</td>
<td>CaCO$_3$</td>
<td>WS</td>
<td>+1.96 ± 0.01</td>
<td>–</td>
<td>MPI-BGC</td>
<td>–</td>
</tr>
<tr>
<td>ALI-J3</td>
<td>Acetanilide</td>
<td>WS</td>
<td>–30.06 ± 0.05</td>
<td>–</td>
<td>MPI-BGC</td>
<td>–</td>
</tr>
<tr>
<td>Cecily</td>
<td>CO$_2$</td>
<td>WS</td>
<td>–3.84 ± 0.01</td>
<td>–</td>
<td>MPI-BGC</td>
<td>–</td>
</tr>
<tr>
<td>Carina-1</td>
<td>Jena air</td>
<td>WS</td>
<td>–82.7 ± 4.0</td>
<td>–47.61 ± 0.09</td>
<td>MPI-BGC</td>
<td>calibration (T. Röckmann, personal communication, 2013)</td>
</tr>
<tr>
<td>Carina-2</td>
<td>Jena air</td>
<td>WS</td>
<td>–85.5 ± 4.0</td>
<td>–47.62 ± 0.12</td>
<td>MPI-BGC</td>
<td>calibration (T. Röckmann, personal communication, 2013)</td>
</tr>
</tbody>
</table>

It is recommended to measure samples against standards with identical material-specific properties (PIT; Werner and Brand, 2001). Under such conditions, measurement artefacts are likely to cancel when, for example, H$_2$O samples are calibrated against H$_2$O standards. However, great care has to be taken when chemically identical or similar CRMs are not available so that sample and standard comprise materials with different chemical properties, which is the case when samples from the 1 mL sample loops into the TC/EA reactor. Typical CH$_4$ flow rates range between 2 and 3 mL min$^{-1}$. For the calibrations of primary CH$_4$ gases, the two sample loops are fed by the same CH$_4$ gas (connecting vent 1 and CH$_4$ port 2). The sample loops are fed by two different gases for the calibration of secondary against primary CH$_4$ gases (Table 1). While CH$_4$ gases are injected manually, H$_2$O is introduced via autosampler.
calibrating CH$_4$ against H$_2$O. Calibration errors may arise when only one or both materials are fractionated during analysis, where the latter is likely to occur with different fractionation factors.

We performed a range of experiments to test for systematic errors during H$_2$O and CH$_4$ analysis. (i) System memory occurs during the isotopic analysis of H$_2$O due to adsorption of H$_2$O onto internal surfaces. System memory is sufficiently minimised by repeated H$_2$O injections and rejection of the first sample in every sequence. Remaining memory effects are corrected for in the evaluation routine as shown by Gehre et al. (2004). System memory is not created by CH$_4$ injections but some $\delta^2$H–CH$_4$ analyses may be affected by desorption of H$_2$O, stemming from previous injections. (ii) We observe a systematic effect of the septum temperature on the resulting $\delta^2$H–H$_2$O and operate the system with a septum temperature of 130°C, where $\delta^2$H–H$_2$O was found stable. $\delta^2$H–CH$_4$ analysis is not affected by septum temperature. (iii) We experimentally optimised the TC/E A reactor temperature and found highest H$_2$ yields, quantitative conversion and hence smallest isotopic fractionation at 1450°C during both $\delta^2$H–H$_2$O and $\delta^2$H–CH$_4$ analyses. Appendix A describes these experiments in greater detail.

The introduction of H$_2$ samples into the ion source of an IRMS leads to the formation of H$^+$ ions that are registered on the HD$^+$ detector, which is accounted for by the so-called “H$_3$-factor correction” (Friedman, 1953; Sessions et al., 2001). The H$_3$-factor correction is experimentally determined and assumed to be constant until re-determined. Determining the H$_3$-factor correction is part of the daily preparation routine at MPI-BGC and shows only minor variation with time. Theoretically, the H$^+$ formation could be dynamic during the experimental period with unknown variability. We matched the H$_2$ peak heights resulting from both CH$_4$ and H$_2$O injections around 5.5 ± 0.5 V in order to minimise the impact of imperfect H$_3$-factor correction. Peak widths ranged around 45 and 60 s for H$_2$O- and CH$_4$-derived H$_2$ peaks respectively. A typical chromatogram of the $\delta^2$H–CH$_4$ calibration including details on peak shape and background is shown in Fig. 3. The similarity between CH$_4$-derived and the H$_2$O-derived H$_2$ peaks allows the use of the standard integration software (ISODAT, Thermo Finnigan, Bremen, Germany).

Megan and Merlin (Table 1) were calibrated in three independent sequences during 3 days against the in-house working standards “WWW-J1” and “BGP-J1” with a wide $\delta^2$H range from −67.0 to −187.1‰ (Table 2). WWW-J1 and BGP-J1 are independently calibrated against international reference waters VSMOW2 and SLAP2 (Table 2). Other CH$_4$ gases were initially also measured against working standards (WWW-J1 and BGP-J1) but were finally calibrated against Megan or Merlin, which were co-analysed in the same measurement sequence in a one-point calibration.

2.3 Referencing pure CH$_4$ for $\delta^{13}$C against LSVEC/MAR-J1 and against other pure CH$_4$ gases

We calibrated $\delta^{13}$C–CH$_4$ in pure CH$_4$ gases after conversion to CO$_2$ using an elemental analyser (EA 1100, CE, Rodano, Italy) coupled to an IRMS (Delta Plus, Thermo Finnigan, Bremen, Germany) via open split (ConFlo III, Thermo Finnigan, Bremen, Germany). This system is routinely used for the analysis of $^{13}$C and $^{15}$N in samples with solid or liquid matrices (Werner et al., 1999; Brooks et al., 2003). We fitted a 1/16 in. tube of 70/30 % Cu/Ni alloy to the EA and used the previously described 10-port valve to inject the CH$_4$ samples into the EA with a 10 mL min$^{-1}$ helium flow (Fig. 4).

The plumbing of the system is designed so that gaseous CH$_4$ and solid CRMs/WSs are applied to the same location inside the combustion reactor of the EA. All samples are combusted at a reactor temperature of 1020°C (Werner et al., 1999) and experience identical analytical treatment thereafter. Following the combustion, each sample passes through a reduction reactor filled with elemental copper, which is kept at 650°C to remove excess O$_2$ and to reduce NO$_x$ if present. The sample is dried by passing through a Naﬁon™ membrane (Perma Pure LLC, Toms River, NJ, USA; not shown in
The 10-port valve for manual CH$_4$ injections is coupled to the EA–IRMS system through custom-made gas inlet into combustion (oxidation) unit for $\delta^{13}$C–CH$_4$ calibration. The proportions of illustrated components are chosen to increase clarity.

Measurement sequences to calibrate primary CH$_4$ gases to the VPDB isotope scale are created by alternating blocks of manual CH$_4$ injections and CRM/WS (Table 2) applications via autosampler. We applied one WS and one CRM (LSVEC) to calibrate the primary CH$_4$ gases in a two-point calibration. While MAR-J1 was used as WS in most experiments, ALI-J1 was used once, during a calibration of Merlin. Megan and Merlin were each calibrated on 3 different days to determine the external reproducibility of the $\delta^{13}$C results. Chromatograms resulting from CH$_4$ and from carbonate analyses using EA–IRMS are displayed in Fig. 5 and show very similar peak shapes for CH$_4$ and carbonates. Typical m/z 44 amplitudes and peak widths were $\sim 7.4 \pm 0.2$ V and 101 $\pm 1$ s for both materials respectively. We connected a primary CH$_4$ and a secondary CH$_4$ gas to the 10-port valve to calibrate the secondary CH$_4$ gases (Table 1) for $\delta^{13}$C in a one-point calibration. All measurement results were corrected for scale compression based on the method suggested in Verkouteren and Klinedinst (2004), using an empirical, mass spectrometer specific correction factor of 1.0056.

2.4 Measurement uncertainty and error propagation

The fully propagated uncertainty for the primary CH$_4$ gases ($U_{pCH4-tot}$) is calculated as

$$U_{pCH4-tot} = \sqrt{u^2_{CRM} + u^2_{WS} + u^2_{pCH4}},$$

where $u_{CRM}$, $u_{WS}$ and $u_{pCH4}$ indicate the uncertainty of the CRM, the applied working standards and the respective primary CH$_4$ gas respectively. Both $u_{WS}$ and $u_{pCH4}$ are calculated as the standard error of the mean of all measurements, multiplied by $t$, Student’s factor for a 95 % confidence limit to account for the limited number of measurements.

The uncertainty for the secondary CH$_4$ gases ($U_{sCH4-tot}$) is then calculated as

$$U_{sCH4-tot} = \sqrt{U^2_{pCH4-tot} + u^2_{sCH4}},$$
where \( u_{\delta^{13}C} \) is the standard error of the mean of all measurements of the respective secondary \( \text{CH}_4 \) gas, multiplied by \( t \), Student’s factor for a 95% confidence limit. Therefore, \( U_{p,\text{CH}_4} \) and \( U_{p,\text{CH}_4} \) indicate the fully propagated uncertainty onto the VPDB or VSMOW isotope scales, representing the traceability chain.

Note that the isotopic composition of LSVEC (Table 2) was recently found to show significant variability, most likely due to adhesion of \( \text{H}_2\text{O} \) and reaction with air-\( \text{CO}_2 \) (e.g. Qi et al., 2016; Schimmelmann et al., 2016). Until this problem is solved, the IAEA, one of the providers of LSVEC, advised to increase the uncertainty of LSVEC, which was hitherto assigned to 0.00%. We follow the recommendation by S. Assonov (Sergey Assonov, IAEA, personal communication, 2016) and Schimmelmann et al. (2016) and adopt an uncertainty of 0.15% for the \( \delta^{13}C \) of LSVEC. Note that the new 0.15% uncertainty of LSVEC represents the largest single contributor to the total uncertainty budget in our \( \delta^{13}C \) calibrations. As a consequence we present the combined uncertainty of the full traceability chain in two versions, the first being the hitherto adopted method using an uncertainty of 0.00% for LSVEC and the second being the method with uncertainty for LSVEC of 0.15%.

### 2.5 Producing synthetic \( \text{CH}_4 \)-in-air standards from pure \( \text{CH}_4 \) and \( \text{CH}_4 \)-free air (JRAS-M16)

The MPI-BGC operates an analytical system (named ARAMIS) to dilute pure \( \text{CO}_2 \) with \( \text{CO}_2 \)-free air to atmospheric \( \text{CO}_2 \) mole fraction without isotopic fractionation (Ghosh et al., 2005). We use ARAMIS to dilute an aliquot of primary or secondary \( \text{CH}_4 \) with \( \text{CH}_4 \)-free air to atmospheric \( \text{CH}_4 \) mole fractions (~2 ppm) in 5 L glass flasks with a final filling pressure of 1.8 bar absolute. The produced synthetic \( \text{CH}_4 \)-in-air standards represent the JRAS-M16 reference gases. The \( \text{CH}_4 \)-free matrix air has been target-mixed from ultra-pure constituents and contains \( \text{N}_2 \), \( \text{O}_2 \), \( \text{N}_2\text{O} \) and Kr at atmospheric levels, so that the composition of the produced \( \text{CH}_4 \)-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 \( \text{CH}_4 \) for \( \delta^{13}C \) (Schmitt et al., 2013). A blank analysis of the \( \text{CH}_4 \)-free air yielded a maximum \( \text{CH}_4 \) blank of 0.5 ppb. Because such a \( \text{CH}_4 \) blank is too small for accurate isotopic analysis on our atmospheric system (Sect. 2.6 and Brand et al., 2016), we choose a mass-balance calculation to determine the maximum blank effect in the synthetic \( \text{CH}_4 \)-in-air standards. Let us assume \( \delta^{2}H \)–\( \text{CH}_4 \) and \( \delta^{13}C \)–\( \text{CH}_4 \) values for the \( \text{CH}_4 \) blank of ~150 and ~40%, respectively, that are typical for fossil \( \text{CH}_4 \). Let us now mix this blank with target \( \text{CH}_4 \) comprising the most depleted \( \delta^{2}H \)–\( \text{CH}_4 \) and \( \delta^{13}C \)–\( \text{CH}_4 \) values we find in our \( \text{CH}_4 \) gas suite with \( \delta^{2}H \) of ~320‰ and \( \delta^{13}C \) of ~70‰. The mixing ratio of blank:target \( \text{CH}_4 \) is 1 : 4000, which reflects the ratio within a synthetic \( \text{CH}_4 \)-in-air mixture with 2 ppm \( \text{CH}_4 \). The maximum blank contribution in this extreme scenario would be 0.04 and 0.007‰ for \( \delta^{2}H \)–\( \text{CH}_4 \) and \( \delta^{13}C \)–\( \text{CH}_4 \), respectively, which is negligible in both cases.

### 2.6 Analytical systems to measure the isotopic composition of \( \text{CH}_4 \) in air at IMAU and MPI-BGC

IMAU: Brass and Röckmann (2010) and Sapart et al. (2011) described the system for the analysis of both \( \delta^{2}H \) and \( \delta^{13}C \) in atmospheric methane at IMAU. \( \text{CH}_4 \) is separated from the other air components by cryogenic traps and gas chromatography before it is converted by either oxidation or pyrolysis for IRMS analysis on \( \text{CO}_2 \) or \( \text{H}_2 \) respectively. For \( \delta^{13}C \) an additional GC column (PoraplotQ, 12.5 m, 0.32 mm ID, Agilent, the Netherlands) was added between the Nafion drying unit and the open split interface in order to remove the interferes from Kr (Schmitt et al., 2013).

MPI-BGC: a new system to measure \( \delta^{2}H \)–\( \text{CH}_4 \) and \( \delta^{13}C \)–\( \text{CH}_4 \) in-air samples was recently developed at the MPI-BGC and is described in greater detail in Brand et al. (2016). The system at MPI-BGC is referred to as iSAAC, in abbreviation for Integrated System for Analysis of Atmospheric Constituents. iSAAC consists of a 16-port sample carousel to take two consecutive 100 mL aliquots of air from a glass flask or high-pressure cylinders for parallel analysis of \( \delta^{2}H \)–\( \text{CH}_4 \) and \( \delta^{13}C \)–\( \text{CH}_4 \), respectively, by continuous-flow GC–IRMS. The two air samples are routed through two identical but independent pre-concentration lines, one for the analysis of \( \delta^{2}H \)–\( \text{CH}_4 \) and one for \( \delta^{13}C \)–\( \text{CH}_4 \). In each line, \( \text{CH}_4 \) is cryogenically separated from the main air constituents in a Hayesep D-filled trap at ~−130°C and cryo-focussed in a further Hayesep D-filled trap at ~−110°C. Each of the two analytical lines is equipped with its own cooling compressor to avoid the use of cryogenic liquids. The separated and cryo-focussed \( \text{CH}_4 \) sample is released into a GC column from where it is routed either through a pyrolysis furnace (kept at 1400°C) to convert the \( \text{CH}_4 \) sample to \( \text{H}_2 \) for \( \delta^{2}H \)–\( \text{CH}_4 \) analysis or through a combustion furnace (kept at 1000°C) to convert the \( \text{CH}_4 \) sample to \( \text{CO}_2 \) for \( \delta^{13}C \)–\( \text{CH}_4 \) analysis. A post-combustion GC column separates the \( \text{CH}_4 \)-derived \( \text{CO}_2 \) from Kr (Schmitt et al., 2013). \( \text{CH}_4 \)-derived \( \text{H}_2 \) and \( \text{CO}_2 \) samples are introduced via open splits into dedicated IRMS instruments, one each for \( \delta^{2}H \)–\( \text{CH}_4 \) and \( \delta^{13}C \)–\( \text{CH}_4 \) analysis. iSAAC has been operational since 2012 to measure air samples with a precision of 1.0 and 0.12‰ for \( \delta^{2}H \)–\( \text{CH}_4 \) and \( \delta^{13}C \)–\( \text{CH}_4 \) respectively. The precision is determined by the performance chart method (Werner and Brand, 2001), determined by the standard deviation (1σ) of all quality control standard measurements, which has been analysed once in every measurement sequence (Brand et al., 2016). The reproducibility of \( \delta^{13}C \)–\( \text{CH}_4 \) analyses ranges around 0.06‰ over the course of a day. All measurements on iSAAC so far have been allocated to the VPDB and VSMOW scales using an in-house WS that was calibrated against “Carina-1” (Table 1).
2.7 Histories to anchor $\delta^2$H–CH$_4$ and $\delta^{13}$C–CH$_4$ to the VSMOW and VPDB scales at IMAU and MPI-BGC

It is the intention of all laboratories analysing $\delta^2$H–CH$_4$ and $\delta^{13}$C–CH$_4$ to reference their samples relative to the VSMOW and VPDB scales respectively. However, possible accuracy errors in the laboratory specific scale anchors often result in inter-laboratory offsets. In order to retrace the potential for calibration offsets between IMAU, MPI-BGC and JRAS-M16, we describe the history of the scale anchors for each laboratory.

IMAU: the calibration strategy at IMAU, including traceability chain and long-term control, is different for $\delta^2$H–CH$_4$ and $\delta^{13}$C–CH$_4$ (Brass and Röckmann, 2010). (i) Three synthetic gas mixtures with CH$_4$ mole fractions of $\sim 9000$ ppm were calibrated for $\delta^2$H–CH$_4$ at the Max Planck Institute for Chemistry (MPI-C) in Mainz, Germany, using a tunable diode laser absorption spectrometer (TDLAS) technique. The TDLAS is described by Bergamaschi et al. (1994) with a measurement precision for $\delta^2$H–CH$_4$ of 5.1 ‰ and an accuracy estimate of similar magnitude. The accuracy estimate is based on a comparison with the calibrations to the VSMOW scale by Dumke et al. (1989), which marks the origin of the isotope scale anchor for $\delta^2$H–CH$_4$ at IMAU. Aliquots of the gases from Bergamaschi et al. (1994) were diluted with synthetic CH$_4$-free air at IMAU to yield reference gases ("Cal1", "Cal2", "Cal3") with the $\delta^2$H–CH$_4$ values initially assigned at MPI-C and atmospheric CH$_4$ levels. Improved measurement precision and inter-laboratory comparisons lead to a $\delta^2$H–CH$_4$ refinement in Cal1, Cal2 and Cal3 with recent values of +21.1, −19.0 and −164.9 ‰ respectively. Cal1, Cal2 and Cal3 represent the primary reference gases for $\delta^2$H–CH$_4$ at IMAU and were used to calibrate the $\delta^2$H–CH$_4$ in the working standard ("SiL") to the VSMOW scale. While Cal2 and Cal3 have become exhausted, Cal1 is still used in regular checks of the calibration scale, together with a set of firn air samples (see ii) that are used for $\delta^{13}$C calibration. (ii) IMAU’s working gas SiL has also been calibrated for $\delta^{13}$C–CH$_4$, this was achieved by co-analysing SiL with a suite of Antarctic firn gas samples, where the $\delta^{13}$C–CH$_4$ of the latter had been determined by two laboratories (MPI-C and the Laboratoire de Géologie et Géophysique de l’Environnement (LGGE), Grenoble, France), using two different techniques (Bräunlich et al., 2001). The $\delta^{13}$C–CH$_4$ scale anchors at LGGE and MPI-C are calibrated at MPI-C against a pure CO$_2$ WS, which itself has been calibrated against NBS 19 (Bergamaschi et al., 2000), which represents the ultimate link to the VPDB scale for the scale anchor at IMAU. Using that method, the suite of firn gas samples was treated as a set of working standards to calibrate SiL to the VPDB scale by propagation from MPI-C and LGGE to IMAU. It is important to note that Brass and Röckmann (2010) highlighted that the firn gas itself is a set of samples and not to be taken for a set of calibration standards. The calibration strategy was revised during 2013 to account for the Kr interference (Schmitt et al., 2013).

MPI-BGC: all measurements on iSAAC use a natural air WS that was calibrated against Carina-1 at MPI-BGC. Carina-1 and Carina-2 are natural air samples that were calibrated for $\delta^2$H–CH$_4$ and $\delta^{13}$C–CH$_4$ at IMAU (Table 2), using the analytical setup described by Brass and Röckmann (2010) and Sapart et al. (2011). While the calibration results of Carina-1 and Carina-2 from IMAU show excellent agreement in CH$_4$ mole fractions (both 1910 ppb), in $\delta^{13}$C–CH$_4$ (within 0.01 ‰), their $\delta^2$H–CH$_4$ values differed by 2.8 ‰ (Table 2). Because both Carina cylinders were filled at the MPI-BGC with Jena air on the same day within a short period of time during stable meteorological conditions, and because their $\delta^{13}$C–CH$_4$ and CH$_4$ mole fractions are in excellent agreement, a true difference in $\delta^2$H–CH$_4$ between the two Carina cylinders seems unlikely. The magnitude of the $\delta^2$H–CH$_4$ offset was smaller than the former $\delta^2$H–CH$_4$ measurement precision at IMAU of ±4 ‰ (Brass and Röckmann, 2010) and was accepted as “agreement within measurement uncertainty” at the time. It is important to note that Carina-1 and Carina-2 were each calibrated on different days and in separate measurement sequences, which does not enable a direct comparison of the two gases. Therefore, a systematic calibration error in one of the two Carina gases is possible. In contrast, the superior measurement precision of iSAAC for $\delta^2$H–CH$_4$ of 1.0 ‰ can resolve a true $\delta^2$H–CH$_4$ difference of 2.8 ‰. However, both Carina-1 and Carina-2 appear indistinguishable in $\delta^2$H–CH$_4$ on iSAAC, as determined during several direct comparisons in independent measurement sequences. Therefore, the $\delta^2$H–CH$_4$ offset between Carina-1 and Carina-2 must be due to an artefact of the calibration at IMAU. Our experiments at MPI-BGC indicate that the calibration of Carina-1 is indeed flawed. The choice to use Carina-1 as scale anchor for all iSAAC measurements at MPI-BGC was made arbitrarily, before it was known that it’s calibration was impacted by an artefact. In hindsight, Carina-2 would have been a better choice as VSMOW scale anchor for $\delta^2$H–CH$_4$ at MPI-BGC. This calibration offset will be furthermore addressed a future comparison with IMAU, where a new system has been developed with an improved precision in for $\delta^2$H–CH$_4$ (Röckmann et al., 2016). All iSAAC measurements are anchored to the VSMOW and VPDB isotope scales based on the described scale propagation from IMAU to MPI-BGC, until JRAS-M16 is established as new m-RM for $\delta^2$H–CH$_4$ and $\delta^{13}$C–CH$_4$ in air.

2.8 Comparison of the existing isotope scales at MPI-BGC with new, synthetic CH$_4$-in-air standards

The synthetic CH$_4$-in-air standards produced in this study (Sect. 2.5) were analysed at MPI-BGC using iSAAC (Sect. 2.6). In that, the synthetic CH$_4$-in-air standards...
are treated as unknown samples and their δ²H–CH₄ and δ¹³C–CH₄ values are determined using Carina-1 as scale anchor (Sect. 2.7). We calculate the isotopic difference (δ_{SAAC}−δ_{pure}) between the measurements on iSAAC and the calibrations of the pure CH₄ gases (Sects. 2.2 and 2.3), which indicates the correction to anchor the measurements at MPI-BGC to JRAS-M16.

2.9 Comparison between CIC and MPI-BGC

Two CH₄ gases, Biogenic and Fossil, were previously calibrated at CIC by Sperlich et al. (2012), who analysed the CH₄-derived CO₂ for δ¹³C–CH₄ by dual-inlet IRMS and the CH₄-derived H₂O for δ²H–CH₄ by either cavity ring-down spectroscopy (CRDS) or TC/EA–IRMS. Sperlich et al. (2012) presented the data with the measurement reproducibility, calculated as the pooled standard deviation of the measurements. Therefore, their uncertainty does not include the uncertainties of the full traceability chain. Furthermore, a statistical provision that accounts for the small number of measurements has not been made by Sperlich et al. (2012). This imposes a hurdle in the comparison with data from MPI-BGC. Therefore, we revise the uncertainty of the CIC data and calculate the full traceability chain as described in Sect. 2.4. Furthermore, all δ¹³C measurements from CIC are affected by a small offset of RM 8563 that has been reported by Coplen et al. (2006b) and are therefore shifted by 0.03 ‰ towards more depleted δ¹³C values. Moreover, the δ¹³C data presented in Sperlich et al. (2012) have not been corrected for scale compression. We are able to correct all CIC data for this effect, because the scale compression factor of the instrument at CIC has been determined (1.0025) at the time the study of Sperlich et al. (2012) was published. Applying the scale compression correction shifts the δ¹³C–CH₄ of Fossil and Biogenic by 0.01 and 0.05 ‰ towards more depleted δ¹³C values respectively. The revised data and uncertainties from CIC and the results from MPI-BGC for Biogenic and Fossil are shown in Table 4 for δ¹³C–CH₄ and in Table 5 for δ²H–CH₄.

We perform two comparisons between CIC and MPI-BGC. (i) The calibration results for Fossil and Biogenic from CIC as published in Sperlich et al. (2012) are compared to the calibrations at MPI-BGC using the methods to calibrate pure CH₄ gases for δ²H–CH₄ and δ¹³C–CH₄ as described in Sects. 2.2 and 2.3. (ii) We performed new combustion experiments at CIC using Fossil and Biogenic and analysed the resulting CO₂ for δ¹³C at both CIC and MPI-BGC. These combustion experiments were made in 2012 but after the publication of Sperlich et al. (2012). Therefore, these experiments provide new data to evaluate the method at CIC. Following the δ¹³C analyses at CIC, the remaining CO₂ gases were cryogenically transferred and flame sealed in glass ampules for δ¹³C analysis at MPI-BGC. The δ¹³C analyses at MPI-BGC were made on “Cora”, a MAT 252 dual-inlet IRMS (Thermo Finnigan, Bremen, Germany) that is used for δ¹³C and δ¹⁸O analysis of CO₂ in air or pure CO₂ gases (Brand et al., 2009b). Unfortunately, the comparison based on the new combustion experiments made at CIC could not include δ²H–CH₄ because the system was not capable to process CH₄ samples large enough to provide sufficient amounts of H₂O.

We use the indices CIC–old for experiments made at CIC and published by Sperlich et al. (2012) and CIC–new for the new combustion experiments at CIC. We use the index MPI–BGC for the analysis at MPI-BGC of CO₂ samples that were combusted at CIC and MPI–BGC for the calibrations of the two CH₄ gases from CIC using the analytical methods at MPI-BGC presented above (Sect. 2.2 and 2.3).

3 Results

3.1 Results for primary CH₄ gas calibrations on the international VSMOW and VPDB isotope scales

We performed 214 repetitive calibration measurements for Megan and Merlin for δ²H–CH₄ and δ¹³C–CH₄, the results are given in Table 3. Megan and Merlin have δ²H–CH₄ values of −168.1 ± 0.7 ‰ and −165.7 ± 0.7 ‰, respectively, and δ¹³C–CH₄ values of −40.76 ± 0.04 and −39.06 ± 0.02 ‰, respectively. Both the δ²H–CH₄ and δ¹³C–CH₄ values are typical for fossil CH₄ (e.g. Quay et al., 1999; Mikaloff Fletcher et al., 2004). The δ¹³C–CH₄ uncertainty in Megan and Merlin increases to 0.16 and 0.15 ‰, respectively, when the suggested uncertainty of 0.15 ‰ for LSVEC is taken into account in the traceability chain (Qi et al., 2016; Schimmelmann et al., 2016). However, we will use the uncertainty budget without the new uncertainty for LSVEC for the evaluation of internal results.

3.2 Results for secondary CH₄ gas calibrations against primary CH₄ gases

We made a total of 260 calibration measurements for the secondary CH₄ gases for δ²H–CH₄ and δ¹³C–CH₄. Altogether, the secondary CH₄ gases cover a large range in δ²H (−320 to +36 ‰) and δ¹³C (−70 to −39 ‰), where the former was achieved by spiking some of the gases with pure CH₃D. The results for secondary CH₄ gas calibrations are shown in Table 3, including the uncertainties of the full traceability chain. We found typical uncertainties on the order of 0.8 ‰ for δ²H–CH₄ calibrations and on the order of 0.07 and 0.17 ‰ for δ¹³C–CH₄ calibrations, where the latter includes the uncertainty of 0.15 ‰ in LSVEC.

3.3 Results from the comparison between CIC and MPI-BGC

Our comparison results for δ¹³C–CH₄ show overall agreement within the uncertainties of the traceability chains (Table 4). The δ¹³C results from the previous and the new com-
Table 3. Results of CH$_4$ isotope calibrations. Gas names as used in main text and their function as primary or secondary CH$_4$ are shown in column 1 and 2 respectively. All uncertainty estimates include the full traceability chain (Sect. 2.4). Note that we provide uncertainty estimates for $\delta^{13}$C–CH$_4$ without and with the uncertainty of 0.15‰ in LSVEC in column 6 and 7 respectively. Martha-1 and Mike-1 were intermittent gases and used to produce Martha-2 and Mike-2.

<table>
<thead>
<tr>
<th>Gas name</th>
<th>Function</th>
<th>$n$ ($\delta^2$H)</th>
<th>$\delta^2$H–CH$_4$ [%ε]</th>
<th>$n$ ($\delta^{13}$C)</th>
<th>$\delta^{13}$C–CH$_4$ [%ε]</th>
<th>$\delta^{13}$C–CH$_4$ [%ε]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Megan</td>
<td>primary</td>
<td>116</td>
<td>$-168.1 \pm 0.7$</td>
<td>15</td>
<td>$-40.76 \pm 0.04$</td>
<td>$-40.76 \pm 0.16$</td>
</tr>
<tr>
<td>Merlin</td>
<td>primary</td>
<td>51</td>
<td>$-165.7 \pm 0.7$</td>
<td>32</td>
<td>$-39.06 \pm 0.02$</td>
<td>$-39.06 \pm 0.15$</td>
</tr>
<tr>
<td>Martha-1</td>
<td>secondary</td>
<td>15</td>
<td>$-176.6 \pm 0.8$</td>
<td>10</td>
<td>$-48.84 \pm 0.07$</td>
<td>$-48.84 \pm 0.17$</td>
</tr>
<tr>
<td>Martha-2</td>
<td>secondary</td>
<td>9</td>
<td>$+36.2 \pm 1.0$</td>
<td>19</td>
<td>$-48.92 \pm 0.06$</td>
<td>$-48.92 \pm 0.16$</td>
</tr>
<tr>
<td>Mike-1</td>
<td>secondary</td>
<td>12</td>
<td>$+44.5 \pm 0.9$</td>
<td>8</td>
<td>$-40.79 \pm 0.09$</td>
<td>$-40.79 \pm 0.17$</td>
</tr>
<tr>
<td>Mike-2</td>
<td>secondary</td>
<td>15</td>
<td>$-80.3 \pm 0.5$</td>
<td>13</td>
<td>$-42.76 \pm 0.05$</td>
<td>$-42.76 \pm 0.16$</td>
</tr>
<tr>
<td>Merida</td>
<td>secondary</td>
<td>12</td>
<td>$-171.7 \pm 0.9$</td>
<td>13</td>
<td>$-60.39 \pm 0.09$</td>
<td>$-60.39 \pm 0.18$</td>
</tr>
<tr>
<td>Melly</td>
<td>secondary</td>
<td>19</td>
<td>$-177.5 \pm 0.7$</td>
<td>15</td>
<td>$-70.04 \pm 0.07$</td>
<td>$-70.04 \pm 0.17$</td>
</tr>
<tr>
<td>Minion</td>
<td>secondary</td>
<td>12</td>
<td>$-182.7 \pm 0.8$</td>
<td>15</td>
<td>$-58.19 \pm 0.05$</td>
<td>$-58.19 \pm 0.16$</td>
</tr>
<tr>
<td>Merkur</td>
<td>secondary</td>
<td>15</td>
<td>$-195.8 \pm 0.9$</td>
<td>19</td>
<td>$-43.05 \pm 0.04$</td>
<td>$-43.03 \pm 0.16$</td>
</tr>
<tr>
<td>Fossil</td>
<td>secondary</td>
<td>15</td>
<td>$-171.9 \pm 0.9$</td>
<td>16</td>
<td>$-39.71 \pm 0.08$</td>
<td>$-39.71 \pm 0.17$</td>
</tr>
<tr>
<td>Biogenic</td>
<td>secondary</td>
<td>25</td>
<td>$-319.8 \pm 0.8$</td>
<td>10</td>
<td>$-56.60 \pm 0.07$</td>
<td>$-56.60 \pm 0.17$</td>
</tr>
</tbody>
</table>

Combustion experiments measured at CIC differ by $-0.03$ and $-0.06\%ε$ ($\delta_{\text{CIC–new}} - \delta_{\text{CIC–old}}$) for Fossil and Biogenic, respectively, which is within the uncertainty of the full traceability chain and furthermore within the system reproducibility as stated in Sperlich et al. (2012). The $\delta^{13}$C differences between the results from the new combustion experiments measured at MPI-BGC* and at CIC ($\delta_{\text{MPI–BGC* – CIC}}$) are $0.10\%$ for Fossil and $0.11\%$ for Biogenic, respectively, and agree well within the combined uncertainty of both methods. Table 4 shows even better agreement for $\delta_{\text{MPI–BGC* – CIC–old}}$. Altogether, the comparisons highlight the reproducibility of CH$_4$ combustion experiments at CIC and the comparability of $\delta^{13}$C measurements in the combustion-derived CO$_2$ at both laboratories.

When comparing the $\delta^{13}$C results from the new calibrations at MPI-BGC to the results based on combustion experiments at CIC, the results from MPI-BGC appear slightly more depleted in $\delta^{13}$C for both Fossil and Biogenic in all comparisons (Table 4). We find the smallest $\delta^{13}$C differences between $\delta_{\text{MPI–BGC}}$ and $\delta_{\text{CIC–new}}$, accounting for $-0.08$ and $-0.09\%ε$ for Fossil and Biogenic respectively. The respective differences increase to $-0.18$ and $-0.20\%ε$ between $\delta_{\text{MPI–BGC}}$ and $\delta_{\text{MPI–BGC*}}$. It is important to note that only the difference found in Biogenic between $\delta_{\text{MPI–BGC}}$ and $\delta_{\text{MPI–BGC*}}$ is outside of the sum of the uncertainties (Table 4). In contrast, we find excellent agreement in all comparisons when the uncertainty of $0.15\%ε$ in LSVEC is taken into account. Table 4 also shows excellent agreement in the determination of the differences between Fossil and Biogenic in all $\delta^{13}$C measurements, which is an important quantity for the evaluation of scale compression.

Comparing the results for $\delta^{2}$H–CH$_4$ between CIC and MPI-BGC shows overall agreement (Table 5). The differences we find in the $\delta^{2}$H–CH$_4$ calibrations between Sperlich et al. (2012) and MPI-BGC ($\delta_{\text{MPI–BGC}}$ – $\delta_{\text{CIC–old}}$) are $-1.8$ and $-2.4\%ε$ for Fossil and Biogenic respectively. Albeit it is slightly larger than the sum of the uncertainties of the measurements at CIC and MPI-BGC for Biogenic, the difference in Fossil is just within the uncertainties of the two methods. Note that the isotopic difference Fossil–Biogenic is homogeneously resolved with $147.9\%ε$ at MPI-BGC and $147.3\%ε$ at CIC respectively.

3.4 Results of $\delta^{2}$H–CH$_4$ and $\delta^{13}$C–CH$_4$ measurements in synthetic CH$_4$-in-air standards to determine compatibility between the propagated isotope scale from IMAU and JRAS-M16 at MPI-BGC

The isotopic difference ($\delta_{\text{SAAC}}$ – $\delta_{\text{pure}}$) is shown in Table 6 and indicates the offset between the scale that was propagated from IMAU to MPI-BGC (Sect. 2.8) and the new synthetic CH$_4$-in-air standards (JRAS-M16), assuming no isotope fractionation during the dilution process. Our experiments show excellent agreement for $\delta^{13}$C–CH$_4$ with an average difference of $+0.03 \pm 0.10\%ε$, thus confirming that the propagated scale from IMAU was already very close to the newly determined scale anchor for $\delta^{13}$C–CH$_4$. For unknown reasons, the $\delta^{13}$C–CH$_4$ measurements of Melly, Fossil and Biogenic show a larger discrepancy between the two methods. Because the discrepancy for Biogenic exceeds the measurement uncertainty by a factor of 3, we have excluded this result from the determination of the laboratory offset. The values for Melly and Fossil are within two uncertainties and are therefore included. For $\delta^{2}$H–CH$_4$ a systematic offset of $+4.2 \pm 1.2\%ε$ is found, confirming that the calibration of Carina-1 and hence the scale propagation from IMAU to MPI-BGC is flawless by an artefact. Obviously,
Carina-2 would have been a closer choice as scale anchor for δ⁵⁷H–CH₄ (Table 2).

4 Discussion

4.1 Discussion on the experimental artefact elimination during δ²H–CH₄ and δ¹³C–CH₄ calibrations in primary and secondary CH₄ gases

We present δ²H and δ¹³C calibrations in pure CH₄ gases against CRMs, WSs and other CH₄ gases. Samples and reference materials were always analysed in the same analytical systems, thereby complying with the PIT as much as possible. The only limitation of the PIT is due to the chemical difference between unknown samples (CH₄) and the known reference materials (carbonates, H₂O) used for anchoring the CH₄ gases to the respective isotope scales. In order to calibrate the primary CH₄ gases accurately, we need to exclude or eliminate material- and method-specific errors (IAEA, 2003), which we discuss in the following.

Quantitative oxidation of CH₄ during δ¹³C–CH₄ analysis requires high reaction temperatures (e.g. Dumke et al., 1989). A major complication during δ¹³C–CH₄ analysis arises when oxidation yields are significantly lower than 100 % (Merritt et al., 1995; Fig. 4 in Sperlich et al., 2012). CH₄ is a potent source of protonation in the IRMS ion source (Anicich, 1993). Introducing unconverted CH₄ together with the CH₄-derived CO₂ sample into the IRMS results in the formation of CO₂H⁺ in the ion source, which produces an isobaric interference on the m/z 45 trace, where the δ¹³C signal is measured. This artefact can be prevented when CO₂ and CH₄ are separated after the oxidation, which we achieve with the post-combustion chromatographic column in both the EA–IRMS system (Sect. 2.3) and iSAAC (Sect. 2.6). Note how this effect would cause an accuracy shift towards more enriched δ¹³C–CH₄ values predominantly during primary CH₄ gas calibrations, because CH₄ samples would be affected by CO₂H⁺ formation in the ion source while the analysis of the used CRMs would not.

We carefully checked the completeness of CH₄ conversion (EA–IRMS and TC/EA–IRMS) by monitoring for residual CH₄ with the IRMS instruments. In the ion source, CH₄ molecules are subject to fragmentation and re-combination processes, resulting in CH₄-typical mass spectra during mass abundance scans in the IRMS (Brunnée and Voshage, 1964). The strongest CH₄-specific signal occurs on the m/z 15 trace (CH₄⁺), which makes the m/z 15 signal a good indicator for incomplete CH₄ conversion (Sperlich et al., 2012). The CH₄⁺ signal at m/z 16 is not suitable for CH₄ quantification due to the interference with the O⁺ signal from CO₂⁺ fragmentation. We tune the m/z 44 collector of the IRMS to monitor the m/z 15 trace during the analysis of a CH₄ sample and find an amplitude of 0.12 mV. From Sperlich et al. (2012) we estimate that about 40 % of the total CH₄ signal in a mass abundance scan is recorded on m/z 15. The total CH₄ signal in the mass abundance scan would therefore amount to ~0.3 mV, which we can compare to the ~7000 mV on m/z 44 from a typical CH₄ injection into the EA–IRMS (e.g. Fig. 5). This approximation suggests a CH₄ oxidation efficiency of >99.9 %.

It has been demonstrated that the introduction of carbonates into the high-temperature oxidation furnace of the EA–IRMS yields a high CO₂ conversion rate and δ¹³C results of high precision and accuracy (Coplen et al., 2006b). In order to test for the completeness of carbonate digestion, we added tungsten trioxide (WO₃) to some of the carbonate samples during weighing (about 1 : 1 by weight). The goal of this experiment is to increase the instantaneous reaction temperature and to provide additional oxygen during the liberation of CO₂ from different carbonates. While the addition of WO₃ had no effect on the analysis of CaCO₃ and Li₂CO₃, it improved the peak shape during BaCO₃ analysis (Table 2).
Table 5. Comparison of $\delta^{2}$H results between CIC and MPI-BGC. Indices of the header are explained in Sect. 2.9 of the main text. The uncertainty of all data includes the full traceability chain (Sect. 2.4), which includes revised uncertainties of the CIC data (Sect. 2.9). The difference Fossil–Biogenic allows us to compare scale compression effects between both methods.

<table>
<thead>
<tr>
<th>Gas name</th>
<th>$\delta^{2}$H$_{\text{CIC-old}}$ [%ε]</th>
<th>$\delta^{2}$H$_{\text{MPI–BGC}}$ [%ε]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fossil</td>
<td>$-170.1 \pm 0.9$</td>
<td>$-171.9 \pm 0.9$</td>
</tr>
<tr>
<td>Biogenic</td>
<td>$-317.4 \pm 0.9$</td>
<td>$-319.8 \pm 0.8$</td>
</tr>
<tr>
<td>Fossil–Biogenic</td>
<td>$147.3$</td>
<td>$147.9$</td>
</tr>
</tbody>
</table>

However, it did not impact on its $\delta^{13}$C. We conclude that the carbonate digestion is not limited by either temperature or oxygen availability and omitted the addition of WO$_3$ in further reactions. Note that the accurate analysis of carbonates is critical for accurate CH$_4$ calibrations, even if CH$_4$ injections themselves are not compromised.

A considerable advantage of the conversion of carbonates in the high-temperature oxidation furnace of the EA–IRMS over other methods (e.g. acid reaction) is that the oxygen isotopic composition is homogenised for all samples. This balances the $\delta^{17}$O correction, which accounts for the isotopic interference between $\delta^{13}$C–CO$_2$ and $\delta^{17}$O–CO$_2$ on $m/z$ 45. The $\delta^{17}$O correction is statistically dependent on the $\delta^{18}$O–CO$_2$ of each individual sample. Hence, any uncertainty arising from the $\delta^{17}$O correction during the calculation of $\delta^{15}$C values from $m/z$ 45 ion currents tends to cancel out. The applied $\delta^{17}$O correction is a function built into the evaluation software of the IRMS. The algorithm and ratio assumptions are based on Assonov and Brenninkmeijer (2001). The same technique had been used to revise the VPDB scale by adding LSVEC as a second scaling point (Coplen et al., 2006b).

The EA–IRMS analysis of carbonates includes a well-characterised blank contribution that is due to the carbon impurities within the tin capsules that are used for carbonate analyses (Werner et al., 1999). In contrast, no such blank is expected when samples are analysed without tin capsules, as would be the case for gaseous CH$_4$ samples. While we did not observe a significant $\delta^{13}$C difference when tin capsules were added to CH$_4$ injections and the $\delta^{13}$C bias was subsequently corrected for or when the $\delta^{13}$C–CH$_4$ analysis was performed without tin capsules. We continuously added the tin capsules to each $\delta^{13}$C–CH$_4$ analysis and applied the routine blank correction to all measurements in compliance with the PIT between analyses of carbonate reference materials and CH$_4$ samples.

For $\delta^{2}$H analyses, we chose an analogue approach and process both H$_2$O and CH$_4$ using the high-temperature reactor of the TC/EA–IRMS system. Possible artefacts can arise mainly from the stronger surface activities of H$_2$O vs. CH$_4$ prior to the conversion to H$_2$ (and CO or carbon). H$_2$O injections can lead to memory effects, which need to be taken into account in $\delta^{2}$H–H$_2$O and subsequent $\delta^{2}$H–CH$_4$ analyses, either by discarding initial injections or making appropriate corrections (Werner and Brand, 2001). H$_2$O injections produced highest H$_2$ yields and stable $\delta^{2}$H–H$_2$O values at reactor temperatures of 1450 °C. Therefore we kept the reactor at 1450 °C during all calibration measurements. In addition, we found a minor dependence of $\delta^{2}$H–H$_2$O on the septum temperature. We experimentally determined a septum temperature of 130 °C at which the effect on $\delta^{2}$H–H$_2$O was insignificant and kept the septum at 130 °C during all calibrations. We describe the experiments on reactor temperature and septum temperature in Appendix A in more detail. Note that it is essential to exclude systematic, material-specific errors to make H$_2$O and CH$_4$ reactions directly comparable for $\delta^{2}$H calibration. Based on these experiments we conclude that the $\delta^{3}$H–CH$_4$ calibrations do not contain measurement errors introduced by bracketing $\delta^{2}$H–H$_2$O analyses.

Table 6. Differences in $\delta^{2}$H–CH$_4$ and $\delta^{13}$C–CH$_4$ between primary/secondary CH$_4$ gas calibrations and ISAAC measurements of the synthetic CH$_4$-in-air standards using the scale anchor based on Carina-1. Differences are calculated as $\delta_{\text{ISAAC}}-\delta_{\text{true}}$. The bottom line shows the average and the standard deviation ($\sigma$) of considered differences, excluding the value of Biogenic (ε) as described in main text.

<table>
<thead>
<tr>
<th>Gas name</th>
<th>$\Delta\delta^{2}$H–CH$_4$ [%ε]</th>
<th>$\Delta\delta^{13}$C–CH$_4$ [%ε]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Megan</td>
<td>$3.9$</td>
<td>$0.05$</td>
</tr>
<tr>
<td>Merlin</td>
<td>$5.6$</td>
<td>$-0.04$</td>
</tr>
<tr>
<td>Minion</td>
<td>$2.7$</td>
<td>$-0.05$</td>
</tr>
<tr>
<td>Melly</td>
<td>$4.3$</td>
<td>$0.13$</td>
</tr>
<tr>
<td>Mike-1</td>
<td>$5.7$</td>
<td>$-0.03$</td>
</tr>
<tr>
<td>Martha-1</td>
<td>$3.2$</td>
<td>$-0.06$</td>
</tr>
<tr>
<td>Fossil</td>
<td>$5.1$</td>
<td>$0.19$</td>
</tr>
<tr>
<td>Biogenic</td>
<td>$3.0$</td>
<td>$0.31$ (ε)</td>
</tr>
<tr>
<td>Average</td>
<td>$+4.2 \pm 1.2$</td>
<td>$+0.03 \pm 0.10$</td>
</tr>
</tbody>
</table>

4.2 Discussion of the comparison between CIC and MPI-BGC

We compare the results of $\delta^{2}$H–CH$_4$ and $\delta^{13}$C–CH$_4$ calibrations achieved by the two independent methods from CIC and MPI-BGC in Tables 4 and 5. Note that the verification of the principle calibration method (MPI-BGC) by an independent method (CIC) is required for the preparation of QCMs when CRMs are not available (IAEA, 2003). The comparison between CIC and MPI-BGC is to some degree representative of the situation of the community analysing atmospheric $\delta^{2}$H–CH$_4$ and $\delta^{13}$C–CH$_4$ without access to international reference air but locally produced or propagated standard gases.

Even though there is no significant difference between the intercomparison results for $\delta^{13}$C–CH$_4$ and the difference in $\delta^{2}$H–CH$_4$ is rather small, there seems to be a systematic pattern that the samples combusted at CIC are generally more...
enriched in both $\delta^2$H and $\delta^{13}$C (Tables 4 and 5). The cause for this offset is not yet fully understood but will be discussed in more detail. The $\delta^{13}$C–CH$_4$ calibrations presented in Table 4 were made on three different IRMS systems with three different working standards. All $\delta^{13}$C measurements were corrected for potential scale compression effects, except from the MPI-BGC$^\ast$ analyses, which were made on an IRMS system specifically tuned to render scale compression effects for $\delta^{13}$C, as demonstrated by Ghosh et al. (2005). Because the difference in $\delta^{13}$C between Fossil and Biogenic is remarkably well resolved in all comparison measurements (Table 4), we conclude that our $\delta^{13}$C comparison does not suffer from a significant scale compression error. Rather, the difference in $\delta^{13}$C between the methods seems related to the method of CH$_4$ conversion. In principle, incomplete CH$_4$ combustion in the experiments at CIC would create a $\delta^{13}$C pattern where the affected experiments appeared more enriched in $\delta^{13}$C. This is because the remaining CH$_4$ fraction in the combustion-derived CO$_2$ gas would be introduced into the dual-inlet IRMS together with the CO$_2$, and form CO$_2$H$^+$ ions, which creates an artefact on $m/z$ 45 (Sect. 4.1). However, we carefully tested every sample for residual CH$_4$ and are confident that the CH$_4$ combustions at CIC have been complete. Therefore, we cannot resolve this difference further.

We also observe a small $\delta^2$H–CH$_4$ offset between CIC and MPI-BGC. The $\delta^2$H measurements at CIC were made using combustion-derived H$_2$O with two different methods (TC/EA–IRMS and CRDS). Moreover, the measurement procedures at CIC included WSs covering the full VSMOW/SLAP scale. In contrast, the direct $\delta^2$H–CH$_4$ analysis of the secondary CH$_4$ gases at MPI-BGC was performed as a one-point calibration against Megan or Merlin with a $\delta^2$H–CH$_4$ similar to that of Fossil (Table 3). Please note that $\delta^2$H scale compression often arises during the analysis of H$_2$O because it interacts with all sorts of surfaces in the analytical system. However, CH$_4$ gas behaves very much like pure H$_2$ in the high-temperature conversion system and a careful $^3$He$^+$-factor determination often results in accurate isotopic distances. If the control of scale compression at MPI-BGC was limited due to the one-point calibration, we would expect the isotopic difference between Biogenic and Fossil to be smaller in the results from MPI-BGC than CIC. This is clearly not the case. The isotopic difference between Biogenic and Fossil ($\delta$Fossil – $\delta$Biogenic) appears to be very similar in the calibrations of both laboratories with 147.3 ‰ at CIC and 147.9 ‰ at MPI-BGC, even showing a slightly larger difference at MPI-BGC (Table 5).

Therefore, we are confident that the observed, small $\delta^2$H offset is not caused by scale compression effects in one of the laboratories. Moreover, the excellent agreement between the experimentally controlled scale compression at CIC and the method at MPI-BGC proves that the analysis at MPI-BGC is free of significant scale compression artefacts over the tested isotopic range of $\sim 150$ ‰.

The comparisons show small differences in the calibration results, but we found no evidence that either one of the two analytical methods is more accurate. Note that the difference in both $\delta^2$H–CH$_4$ and $\delta^{13}$C–CH$_4$ exceeds the compatibility goal of 1 and 0.02 ‰ by a factor of 2 to 10 respectively (WMO, 2014). We interpret the results of this comparison to reflect calibration differences between laboratories that are to be expected, when CRMs are not available. Finally, we conclude that our new method is as capable to calibrate CH$_4$ gases to the international isotope scales and that it is as accurate as the method presented by Sperlich et al. (2012). However, we think that our new methods are more suitable for the task to produce and maintain a suite of calibration gases for the following reasons.

- The methods at MPI-BGC are more time efficient than the method of Sperlich et al. (2012). While the new methods at MPI-BGC can be used to calibrate an entire suite of CH$_4$ gases within a relatively short time, the method of Sperlich et al. (2012) is capable of processing only one sample per day.

- The new MPI-BGC methods are based on continuous-flow IRMS and follow the PIT to the highest possible degree. In comparison, the method of Sperlich et al. (2012) is based on the combustion of CH$_4$ in an offline reactor, which requires re-oxidation after every sample and partial dismantling of the system to retrieve the sample for isotopic analysis. Because the analytical system at CIC could theoretically be at a different state for every sample (oxidation state, air leak rate) and because the system at CIC does not allow us to compare two CH$_4$ gases directly against each other, the methods at MPI-BGC are superior in the ability to fulfill the PIT. Even though the method at CIC proved to be very reproducible, we cannot rule out that a variation in the oxidation state of the reactor or an undetected air leakage into the system would affect the analysis of some CH$_4$ samples more than others. Because fulfilling the PIT is of paramount importance for isotope ratio analysis (e.g. Werner and Brand, 2001; Schimmelmann et al., 2016), we believe the method at MPI-BGC is less vulnerable to measurement errors in future calibrations.

4.3 Discussion on the compatibility between the scale anchors for $\delta^2$H–CH$_4$ and $\delta^{13}$C–CH$_4$ as propagated from IMAU to MPI-BGC and JRAS-M16

We interpret the excellent agreement between the $\delta^{13}$C and CH$_4$ calibrations in Carina-1 and Carina-2 from IMAU (Table 2) that both gases are precisely referenced and suitable for scale propagation from IMAU to MPI-BGC. The synthetic CH$_4$-in-air standards were analysed on iSAAC for $\delta^2$H–CH$_4$ and $\delta^{13}$C–CH$_4$ and their isotope values were assigned using a WS that was calibrated against Carina-1. We can then
interpret the $\delta^{13}C$ difference between the iSAAC measurement and the calibrated synthetic CH$_4$-in-air standards of $+0.03 \pm 0.10\%_e$ as an accurate estimate for the calibration offset between the propagated scale anchor at MPI-BGC and the newly developed JRAS-M16.

Unfortunately, the situation is currently less straightforward for $\delta^2H$–CH$_4$. The two WSs Carina-1 and Carina-2 were calibrated at IMAU with a difference in $\delta^2H$–CH$_4$ of $2.8\%_e$ that was insignificant at the time (Table 2). Because Carina-1 and Carina-2 appear indistinguishable in $\delta^2H$–CH$_4$ when compared to iSAAC with a measurement precision for $\delta^2H$–CH$_4$ of $1.0\%_e$ (Sect. 2.6), we cannot determine the laboratory offset with the same certainty as for $\delta^{13}C$–CH$_4$. If either Carina-1 or Carina-2 were representative for the calibrations at IMAU, the $\delta^2H$–CH$_4$ offset between the laboratories would amount to $+4.2 \pm 1.2$ or $+1.4 \pm 1.2\%_e$ respectively. A further comparison that includes new measurements on the current system at IMAU is required to determine the offset $\delta^2H$–CH$_4$ accurately. This offset can be resolved, for example, when a set of synthetic CH$_4$-in-air standards (JRAS-M16) is analysed at IMAU in future.

4.4 Discussion on possible use of synthetic CH$_4$-in-air standards in future

We demonstrated the ability to test the compatibility between IMAU and MPI-BGC by comparing scale anchors that were previously propagated from IMAU to MPI-BGC to JRAS-M16 gases. Future developments include an inter-laboratory comparison to test whether a dedicated set of our synthetic CH$_4$-in-air standards (JRAS-M16) could provide a community anchor to the VPDB and VSMOW scales with documented accuracy. A further important test would be to determine to what extent the use of centrally calibrated standard gases could increase compatibility. A recent incidence provides a good example for the vulnerability of $\delta^{13}C$–CH$_4$ observations in the atmosphere without suitable m-RM.

LSVEC, the second CRM anchor to the VPDB scale, has recently been discovered to be less reliable than anticipated. Until further notice, LSVEC is suggested to be treated with an enhanced $\delta^{13}C$ uncertainty of $0.15\%_e$ (S. Assonov, personal communication, 2016). It is important to appreciate that this uncertainty is fully added to the uncertainty of $\delta^{13}C$–CH$_4$ measurements, due to the similarity of LSVEC ($-46.6\%_e$) and tropospheric CH$_4$ ($-47.5\%_e$) in $\delta^{13}C$. That is, the new uncertainty of LSVEC contributes the largest component in the full error budget of $\delta^{13}C$–CH$_4$ analysis. Note that the suggested uncertainty of LSVEC is (i) on the order of the seasonal $\delta^{13}C$–CH$_4$ cycle in the Southern Hemisphere and (ii) a multiple of the analytical precision of laboratories monitoring $\delta^{13}C$–CH$_4$. If measurements of $\delta^{13}C$–CH$_4$ considered the new uncertainty for LSVEC, the significance of signals such as the seasonal variability in the Southern Hemisphere would be lost on the cost of a better representation of accuracy. Including the uncertainty of LSVEC may further impact on the compatibility between several laboratories and, for example, suggest an artificially imposed spatial $\delta^{13}C$–CH$_4$ gradient, based on calibration artefacts. We advocate the scientific gain when accuracy and compatibility are differentiated (WMO, 2014). The community benefits from a referencing method that enables a compatibility level that is smaller than the atmospheric $\delta^{13}C$–CH$_4$ signal to resolve spatiotemporal $\delta^{13}C$–CH$_4$ differences as primary goal. We think that establishing JRAS-M16 as a community scale anchor could be a valuable step towards reaching this goal. As appropriate for any scale anchor that is intended to be usable for the whole community over long periods of time, the scale anchors will have to be re-calibrated frequently in order to detect possible drifts or to improve and correct previous assignments. The results of these efforts will be made available to the public at regular intervals.

We propose the distribution of JRAS-M16, a set of synthetic CH$_4$-in-air standards in 5 L glass flasks. While two JRAS-M16 gases shall be used as calibration standard, an optional third JRAS-M16 gas can be used as unknown that is calibrated against the known JRAS-M16 gases as measurement control standard. This experiment would simulate the case when all participating laboratories measure the same sample directly against the same m-RM using the method that is otherwise applied to every sample in the respective laboratory and has the potential to determine the achievable compatibility. A further possibility to share the JRAS-M16 scale anchor would be to send cylinders with air-WSs to MPI-BGC for calibration. Because a dedicated target of this work is to achieve best possible accuracy with JRAS-M16, we provide the uncertainty of the full traceability chain. Once a new CRM has been found in replacement of LSVEC, the $\delta^{13}C$–CH$_4$ and the traceability chain of JRAS-M16 will be revised accordingly. This will also be made upon future CRM revisions or replacements.

5 Conclusions

The number of laboratories that measure isotope ratios of atmospheric CH$_4$ is growing and combining data from multiple laboratories could enable new science and increasingly powerful analysis. 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. To overcome this problem and to improve compatibility between laboratories, we produced synthetic CH$_4$-in-air standards (JRAS-M16). We modified standard online IRMS techniques to calibrate pure CH$_4$ gases for $\delta^2H$ and $\delta^{13}C$ on international VSMOW and VPDB isotope scales respectively. Because such instrumentation is available to many isotope laboratories, our technical modifications and experiments can be reproduced elsewhere. Eight of the calibrated CH$_4$ gases were diluted with CH$_4$-free air in 5 L glass flasks to produce synthetic CH$_4$-in-air standards.
with known $\delta^{2}H$–CH$_{4}$ and $\delta^{13}C$–CH$_{4}$ values. These synthetic gas mixtures were then analysed on a newly developed system (iSAAC) to measure $\delta^{2}H$–CH$_{4}$ and $\delta^{13}C$–CH$_{4}$ in air samples. Hitherto, iSAAC used working standards as scale anchors for $\delta^{2}H$–CH$_{4}$ and $\delta^{13}C$–CH$_{4}$, which were calibrated at a partnering institute (IMAU). The history of the propagated isotope scales goes more than 2 decades back in time and includes the propagation between several laboratories. We determine $\delta^{2}H$–CH$_{4}$ and $\delta^{13}C$–CH$_{4}$ in our synthetic CH$_{4}$-in-air standards using the scale anchor propagation from IMAU and compare the results with our calibration results for $\delta^{2}H$–CH$_{4}$ and $\delta^{13}C$–CH$_{4}$. We use this method to determine the $\delta^{13}C$–CH$_{4}$ offsets between the scale anchor that was propagated from IMAU and JRAS-M16, thereby providing a method to improve compatibility. Further comparisons are required to determine the offset for $\delta^{2}H$–CH$_{4}$.

We welcome other laboratories to further test our calibrations by analysing JRAS-M16 air sets, which will be available upon request. Another possibility could be to have cylinders with air-WSs sent to MPI-BGC for calibration using JRAS-M16 as scale anchor. JRAS-M16 may help laboratories to anchor $\delta^{2}H$–CH$_{4}$ and $\delta^{13}C$–CH$_{4}$ observations to unified community scale anchors. This might be a useful step towards reaching the compatibility goals between laboratories, leading to an improved understanding of atmospheric CH$_{4}$.

Future work includes a revision of the $\delta^{13}C$–CH$_{4}$ calibrations once the replacement for LSVEC is established. This will reduce the uncertainty of the $\delta^{13}C$–CH$_{4}$ scale anchors significantly. The LSVEC replacement should extend to the $\delta^{13}C$-depleted range of biogenic CH$_{4}$ gases.

6 Data availability

The results of our final calibrations with the associated uncertainties of the full traceability chains are published as a Supplement to this paper. The supplementary data file also contains the revised calibrations of the data by Sperlich et al. (2012). These include corrections for the offset in RM8563 and for scale compression effect in the IRMS at CIC.
Appendix A: Experiments to enhance the performance of the analytical system for the calibration of $\delta^2$H–CH$_4$ with H$_2$O

**Figure A1.** The $\delta^2$H variation of H$_2$O injections with septum temperatures. Blue circles show average $\delta^2$H–H$_2$O values for septum temperatures above 90°C, the black line is the quadratic polynomial fit to the data above 90°C while red diamonds display $\delta^2$H–H$_2$O values at septum temperatures below 90°C. The error bars show 1σ standard deviations and the grey-dashed lines indicate the typical precision limit of 1 ‰ for $\delta^2$H–H$_2$O around the $\delta^2$H–H$_2$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 $\delta^2$H–H$_2$O analyses remain within a typical precision level as long as the septum temperature is controlled to $\sim 130 \pm 10$ °C.

The injection of H$_2$O samples into the reactor is critical because it is prone to isotopic fractionation (Werner and Brand, 2001). This fractionation is mainly caused by system memory due to adhesion of injected H$_2$O to the reactor walls. The isotopic fractionation can be overcome by repetitive injections of H$_2$O samples with identical isotopic composition, thereby overwriting the memory effect until it reaches a marginal level. For H$_2$O analyses under constant analytical conditions (e.g. constant reactor temperature), the adhesion effect is a function mainly of the amount of injected H$_2$O sample. Moreover, the effect on the isotopic composition scales with the isotopic difference between two consecutive samples (Gehre et al., 2004). Because there is no adhesion of the sample during CH$_4$ analysis, this memory effect is most pronounced only during the analysis of H$_2$O in our study. Subsequent CH$_4$ analysis does not contribute to system memory but can still be affected by H$_2$O desorption from internal surfaces of the analytical system. Therefore, memory effects of H$_2$O can propagate into the CH$_4$ calibrations. Memory effects are identified in a series of replicate H$_2$O measurements and are corrected for by modelling the memory function as described in Gehre et al. (2004) and Brand et al. (2009a) on a routine basis, as our system has been used for isotope analysis of H$_2$O samples for more than a decade. We conclude that our results are free of artefacts arising from sample memory.

Isotopic fractionation during the analysis of the reference waters can also be caused by insufficiently heated septa (Gehre et al., 2004). We injected 106 identical H$_2$O samples while we increased the septum temperatures in nine steps from 76 to 137°C. In general, we observed a $\delta^2$H enrichment with increasing septum temperature. A systematic increase of $\delta^2$H–H$_2$O with septum temperature is apparent above 90°C until $\delta^2$H–H$_2$O values plateau at septum temperatures around 130°C (Fig. A1, blue circles). The stabilising $\delta^2$H–H$_2$O at high temperatures suggests quantitative H$_2$O processing without significant isotope fractionation, in line with previous observations (Gehre et al., 2004). In contrast, the three $\delta^2$H–H$_2$O values below 90°C (red diamonds) show an insignificant but slight increase in $\delta^2$H–H$_2$O with septum temperature, which deviated from the pattern above 90°C. We cannot explain the mismatch between the two patterns above and below 90°C. We speculate that the initial heating of the septum to temperatures between 70 and 90°C caused the desorption of accumulated of H$_2$O, which was desorbed once the septum was heated to temperatures above 90°C.

Quantitative conversion of both CH$_4$ and H$_2$O in the high-temperature reactor is of utmost importance for our study, because incomplete conversion causes isotopic fractionation in the reaction products (e.g. Burgoine and Hayes, 1998; Hilkert et al., 1999; Gehre et al., 2004). The reactor temper-
ature is critical for the efficiency of the conversion process. We performed an experiment with CH$_4$ and H$_2$O injections at different reactor temperatures (Fig. A2). For water injections we observe a pronounced, nonlinear $\delta^2$H–H$_2$O change of $\sim 15$‰ with reactor temperature increase from 1300 to 1450°C, reaching a plateau above 1400°C. The pattern is consistent with previous observations in both trend and magnitude (Gehre et al., 2004). In contrast, the linear fit for $\delta^2$H–CH$_4$ increases by only about 1‰ over the 150°C temperature range. However, the slope is statistically insignificant as shown by the 95% confidence interval of the linear fit (Fig. A2). This analyte-specific isotope variation is also reflected in the areas of the H$_2$O and CH$_4$-derived H$_2$ peaks (Fig. A2) (with some significant scatter in the data). While the H$_2$O-derived H$_2$ peak areas increase with increasing reactor temperature, the CH$_4$-derived H$_2$ peak areas remain constant within the error bars throughout the experiments. For an unknown reason, three out of six H$_2$ peaks that resulted from H$_2$O injections at 1400°C were by 10–15 standard deviations smaller than the remaining three peaks. We present the averages and 1σ standard of the H$_2$ peaks with and without removal of these outliers in Fig. A2, which shows the exceptional pattern at 1400°C. Despite this peak size variability, the isotopic composition of all H$_2$O injections at 1400°C is in good agreement. Our experiments indicate that reactor temperatures in excess of 1400°C are required especially for quantitative conversion of H$_2$O, while the effects of reactor temperature on both yield and the isotopic composition of CH$_4$-derived H$_2$ are comparably small. Therefore, we operate the reactor at a temperature of 1450°C to guarantee quantitative conversion without isotope fractionation of both H$_2$O (Gehre et al., 2004) and CH$_4$ (Burgoyne and Hayes, 1998; Hilkert et al., 1999).
The Supplement related to this article is available online at doi:10.5194/amt-9-3717-2016-supplement.

Acknowledgements. Financial support was provided within the European Commission projects InGOS and IMECC. N. A. M. Uitslager’s visit in Jena was made possible by the Erasmus program. Support by Huilin Chen and Harro Meijer (University of Groningen) is gratefully appreciated. We are indebted to Ingeborg Levin for making archived air samples from Neumayer station available for comparison measurements and to Gordon Brailsford and Sara E. Mikaloff Fletcher for valuable comments during the preparation of the manuscript. We are thankful for the insightful reviews of an anonymous reviewer and Ingeborg Levin as well as the comments provided by Sergey Assonov, which altogether helped to improve this paper.

The article processing charges for this open-access publication were covered by the Max Planck Society.

Edited by: F. Keppler
Reviewed by: I. Levin and one anonymous referee

References


