Interlaboratory comparison of $\delta^{13}C$ and $\delta^D$ measurements of atmospheric $\text{CH}_4$ for combined use of data sets from different laboratories

Taku Umezawa$^{1,2}$, Carl A. M. Brenninkmeijer$^1$, Thomas Röckmann$^3$, Carina van der Veen$^3$, Stanley C. Tyler$^{4,5}$, Ryo Fujita$^6$, Shinji Morimoto$^{6,7}$, Shuji Aoki$^6$, Todd Sowers$^8$, Jochen Schmitt$^9$, Michael Bock$^9$, Jonas Beck$^9$, Hubertus Fischer$^9$, Sylvia E. Michel$^{10}$, Bruce H. Vaughn$^{10}$, John B. Miller$^{10}$, James W. C. White$^{10}$, Gordon Brailsford$^{11}$, Hinrich Schaefer$^{11}$, Peter Sperlich$^{11}$, Willi A. Brand$^{12}$, Michael Rothé$^{12}$, Thomas Blunier$^{13}$, David Lowry$^{14}$, Rebecca E. Fisher$^{14}$, Euan G. Nisbet$^{14}$, Andrew L. Rice$^{15}$, Peter Bergamaschi$^{16}$, Cordelia Veidt$^{17}$, and Ingeborg Levin$^{17}$

1Max Planck Institute for Chemistry, Atmospheric Chemistry Department, 55128 Mainz, Germany
2Center for Environmental Measurement and Analysis, National Institute for Environmental Studies, Tsukuba 305-8506, Japan
3Institute for Marine and Atmospheric research Utrecht, Utrecht University, Utrecht, the Netherlands
4Earth System Science Department, University of California, Irvine, USA
5Chemistry Department, Norco College, Norco, CA 92860, USA
6Center for Atmospheric and Oceanic Studies, Graduate School of Science, Tohoku University, Sendai, Japan
7National Institute of Polar Research, Tokyo, Japan
8The Earth and Environmental Systems Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA
9Climate and Environmental Physics, Physics Institute and Oeschger Center for Climate Change Research, University of Bern, 3012 Bern, Switzerland
10Institute of Arctic and Alpine Research, University of Colorado Boulder, Boulder, Colorado, USA
11National Institute of Water and Atmospheric Research, Wellington 6021, New Zealand
12Max Planck Institute for Biogeochemistry, 07745 Jena, Germany
13Centre for Ice and Climate, University of Copenhagen, Copenhagen, Denmark
14Department of Earth Sciences, Royal Holloway, University of London, Egham, UK
15Department of Physics, Portland State University, Portland, OR 97207, USA
16European Commission Joint Research Centre, Institute for Environment and Sustainability, Ispra (Va), Italy
17Institute for Environmental Physics, Heidelberg University, 69120 Heidelberg, Germany

Correspondence: Taku Umezawa (umezawa.taku@nies.go.jp)

Received: 31 July 2017 – Discussion started: 9 August 2017
Revised: 16 January 2018 – Accepted: 17 January 2018 – Published: 2 March 2018

Abstract. We report results from a worldwide interlaboratory comparison of samples among laboratories that measure (or measured) stable carbon and hydrogen isotope ratios of atmospheric CH$_4$ ($\delta^{13}C$-CH$_4$ and $\delta^D$-CH$_4$). The offsets among the laboratories are larger than the measurement reproducibility of individual laboratories. To disentangle plausible measurement offsets, we evaluated and critically assessed a large number of intercomparison results, some of which have been documented previously in the literature. The results indicate significant offsets of $\delta^{13}C$-CH$_4$ and $\delta^D$-CH$_4$ measurements among data sets reported from different laboratories; the differences among laboratories at modern atmospheric CH$_4$ level spread over ranges of 0.5 ‰ for $\delta^{13}C$-CH$_4$ and 13 ‰ for $\delta^D$-CH$_4$. The intercomparison results summarized in this study may be of help in future attempts to harmonize $\delta^{13}C$-CH$_4$ and $\delta^D$-CH$_4$ data sets from
different laboratories in order to jointly incorporate them into modelling studies. However, establishing a merged data set, which includes $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ data from multiple laboratories with desirable compatibility, is still challenging due to differences among laboratories in instrument settings, correction methods, traceability to reference materials and long-term data management. Further efforts are needed to identify causes of the interlaboratory measurement offsets and to decrease those to move towards the best use of available $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ data sets.

1 Introduction

Methane (CH$_4$) is an important anthropogenic and natural greenhouse gas, and it also has a large role in atmospheric chemistry through its reaction with the hydroxyl radical. Since individual CH$_4$ source types have characteristic isotopic signatures and loss processes are associated with specific kinetic isotope effects, carbon and hydrogen isotope ratios of CH$_4$ ($\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$) have been useful in constraining the global CH$_4$ budget. Dictated by global mass balance, the average isotopic composition of CH$_4$ in the atmosphere ($\delta^{13}$C-CH$_4$ or $\delta$D-CH$_4$) equals the flux-weighted isotopic composition of the sources, corrected for the total kinetic isotope effects of removal processes (e.g. Stevens and Rust, 1982; Cicerone and Oremland, 1988; Quay et al., 1991, 1999; Miller et al., 2002; Turner et al., 2017; Rigby et al., 2017). It has been pointed out that assignment of representative isotopic signatures of various CH$_4$ sources remains uncertain due to their large spatial and temporal variability across the globe (e.g. Sherwood et al., 2017), which could result in large uncertainties of isotope-based estimates of the global CH$_4$ budget (Schwietzke et al., 2016). Nonetheless, the value of isotope measurements was amply demonstrated by recent studies which suggested shifts in the global CH$_4$ source over the last decades (Schaefer et al., 2016; Rice et al., 2016; Nisbet et al., 2016; Schwietzke et al., 2016); without isotopic analyses such conclusions would have been difficult to achieve. The isotopic ratios are commonly reported using the delta notation:

$$\delta = \frac{R_{\text{sample}}}{R_{\text{standard}}} - 1,$$

where $R$ represents the atomic ratio of the less abundant isotope over the most abundant isotope in the sample or the standard. Conventionally, measured values are reported relative to the international isotope-scale VPDB (Vista Peak Dee Belemnite) for $\delta^{13}$C-CH$_4$ and VSMOW (Vienna Standard Mean Ocean Water) for $\delta$D-CH$_4$ in per mil.

Given that the atmospheric lifetime of CH$_4$ is about a decade, its variation in background air is relatively small. For that reason, its mole fraction and isotopic measurements have to have high precision and accuracy. For $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$, researchers have achieved measurement reproducibility of $<0.1$ for $\delta^{13}$C-CH$_4$ and $<2\%$ for $\delta$D-CH$_4$. Incorporating $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ data sets in chemistry transport models is useful for quantitatively separating different CH$_4$ source categories and attempts have been made to reduce uncertainties in the global CH$_4$ budget (e.g. Fung et al., 1991; Hein et al., 1997; Mikaloff Fletcher et al., 2004a, b; Monteil et al., 2011; Kirschke et al., 2013; Ghosh et al., 2015; Rice et al., 2016; Schaefer et al., 2016; Schwietzke et al., 2016; Röckmann et al., 2016; Turner et al., 2017; Rigby et al., 2017). However, although an increasing number of $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ data have been reported over the last decades, significant measurement offsets among laboratories have been found for both $\delta^{13}$C-CH$_4$ (e.g. Levin et al., 2012) and $\delta$D-CH$_4$ (Bock et al., 2014). It is clear that both traceability to the standard scales and interlaboratory comparisons (intercomparisons) are indispensable for combined use of $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ data from different laboratories. Many such intercomparisons have already been made, either on an ad hoc basis or on a more organized scale. However, a systematic evaluation of the underlying calibrations and related measurement offsets among laboratories has been lacking. It is also noted that some measurement programmes for $\delta^{13}$C-CH$_4$ and/or $\delta$D-CH$_4$ have been discontinued, and maintaining access to such data sets including well-established interlaboratory offsets is important.

Here we combine and evaluate the existing comparison results to quantify interlaboratory measurement differences in order to facilitate the use of $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ data. This study therefore opens the possibility for merging historic CH$_4$ isotope data reported from multiple laboratories (i.e. synthesis analysis of the existing data sets) for a better understanding of the global CH$_4$ budget.

We first present a technical overview of atmospheric $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ measurements and potential causes of measurement offsets among currently available data sets (Sect. 2), and then we summarize measurement methods by the laboratories that have conducted $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ measurements for air and ice core samples (Sect. 3). In Sect. 4, we report new intercomparison exercises between some groups. We then link the intercomparison results through a survey of previously published intercomparisons and provide the current best estimates of measurement offsets among data sets from different laboratories (Sect. 5). Finally, we summarize the current status and briefly discuss possible causes of the measurement offsets as well as remaining issues that should be kept in mind when combining the use of currently existing data sets of isotopic composition of CH$_4$ (Sect. 6).
Table 1. List of laboratories that conduct measurements of $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$. For each laboratory, measurement systems and relevant information that could have contributed, the interlaboratory measurement offsets are summarized. Brackets in the RM column indicate the laboratory from which the original standard scale was propagated. See Fig. 1 for overview of the past intercomparison exercises, Fig. 2 for intercomparison summary and the list of participating institution/project acronyms in the text for the laboratory names.

<table>
<thead>
<tr>
<th>No.</th>
<th>Lab</th>
<th>System</th>
<th>IRMS</th>
<th>$^{12}$O correction$^a$</th>
<th>Kr interface</th>
<th>Additional correction</th>
<th>RM</th>
<th>Reference$^c$</th>
<th>$\delta^{13}$C-CH$_4$ offset with respect to NIWA</th>
<th>$\delta$D-CH$_4$ offset with respect to IMAU</th>
<th>Data availability$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NIWA</td>
<td>DI</td>
<td>C1/C2</td>
<td>N</td>
<td>–</td>
<td>N</td>
<td>NBS-19, IAEA-C0-9, LSVEC</td>
<td>Not measured</td>
<td>R1, R2, R3, R4, R5, R6, R7, R8, R9, R10, R11, R12, R13</td>
<td>–</td>
<td>Not measured</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>BMAU</td>
<td>CF</td>
<td>C2</td>
<td>Y</td>
<td>PCS</td>
<td></td>
<td>NBS-19 (MPC) VSMOW &amp; SLAP (MPC)</td>
<td>R18</td>
<td>R10, R12, R18, R19, R20, R21, R22, R23</td>
<td>–0.04 ± 0.07</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>MPIC</td>
<td>DI &amp; TDLAS</td>
<td>MAT 252</td>
<td>C2</td>
<td>N</td>
<td>N</td>
<td>NBS-19 VSMOW &amp; SLAP</td>
<td>R27</td>
<td>R18, R19, R24</td>
<td>–0.04 ± 0.07</td>
<td>+0.3 ± 1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>MP3-BGC</td>
<td>CF</td>
<td>C4</td>
<td>Y</td>
<td>PCS</td>
<td></td>
<td>NBS-19 VSMOW 2 &amp; SLAP 2</td>
<td>R29</td>
<td>R21, R30, R21, R20, R22, R23</td>
<td>–0.07 ± 0.12</td>
<td>–4.2 ± 1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>UCI</td>
<td>DI</td>
<td>MAT 252</td>
<td>C2</td>
<td>N</td>
<td>N</td>
<td>NBS-19 &amp; IAEA-CO-9 Oitech Gas</td>
<td>R7, R31</td>
<td>R7, R12, R15, R22, R32, R33</td>
<td>–0.01 ± 0.09</td>
<td>–10.7 ± 0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>TU</td>
<td>CF</td>
<td>C3</td>
<td>Y</td>
<td>No</td>
<td></td>
<td>NBS-19 VSMOW &amp; SLAP</td>
<td>R37</td>
<td>R23, R37, R23, R37</td>
<td>+0.36 ± 0.08</td>
<td>–13.1 ± 0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>NIPR</td>
<td>CF</td>
<td>MAT 252</td>
<td>C3</td>
<td>N</td>
<td>N</td>
<td>NBS-19 (TU) Not measured</td>
<td>R8</td>
<td>R8, R23, R37</td>
<td>+0.33 ± 0.04</td>
<td>Not measured</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>UW</td>
<td>DI</td>
<td>MAT 251</td>
<td>Information not available</td>
<td>N</td>
<td>–</td>
<td>NBS-19 VSMOW &amp; SLAP</td>
<td>R3</td>
<td>R2, R3, R4, R32</td>
<td>+0.02 ± 0.14</td>
<td>No comparison available</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>UHEI</td>
<td>DI</td>
<td>MAT 252</td>
<td>C2 with coefficients $a=0.5$ &amp; $K=0.008335$</td>
<td>N</td>
<td>–</td>
<td>NBS-19 VSMOW &amp; SLAP</td>
<td>R5</td>
<td>R4, R5, R6, R19, R30</td>
<td>–0.17 ± 0.03</td>
<td>–3.8 ± 1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>INSTAAR</td>
<td>CF</td>
<td>Optima/isoprime</td>
<td>C3</td>
<td>Y</td>
<td>PCS</td>
<td>NBS-19 &amp; IAEA-CO-9 (3C2)</td>
<td>R15</td>
<td>R4, R7, R9, R11, R12, R13, R15, R19, R41</td>
<td>+0.13 ± 0.02</td>
<td>–13.2 ± 1.3</td>
</tr>
</tbody>
</table>
Table 1. Interlaboratory comparison of \( \delta^{13}C \) and \( \delta^{15}D \) measurements of CH\(_4\).

<table>
<thead>
<tr>
<th>No</th>
<th>Lab</th>
<th>System</th>
<th>IRMS</th>
<th>O correction</th>
<th>Kr interference</th>
<th>Additional correction</th>
<th>Reference</th>
<th>RM</th>
<th>( \delta^{13}C )-CH(_4) offset with respect to NIWA</th>
<th>( \delta^{15}D )-CH(_4) offset with respect to IMAU</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>RHUL CF</td>
<td>DI</td>
<td>Isoprime</td>
<td>C3</td>
<td>N – Daily offset with respect to working</td>
<td>NBS-19 &amp; IAEA-CO-9</td>
<td>Not measured</td>
<td>R43, R42</td>
<td>R9, R11, R13</td>
<td>Not measured</td>
</tr>
<tr>
<td>12</td>
<td>PDX CF</td>
<td>Delta V C2</td>
<td>Y</td>
<td>DI</td>
<td>PANGAEA</td>
<td>R33, R34</td>
<td>R32, R33</td>
<td>+0.08 ± 0.03</td>
<td>Not measured</td>
<td>CEDA &amp; On request to Euan G. Nisbet (R11)</td>
</tr>
<tr>
<td>13</td>
<td>PSU CF</td>
<td>MAT 252</td>
<td>C2</td>
<td>Y</td>
<td>Raw ion current correction/DI</td>
<td>VSMOW &amp; SLAP (UHEI)</td>
<td>R20, R44</td>
<td>R10, R12, R19, R20</td>
<td>R12, R20, R24</td>
<td>+0.03 ± 0.16</td>
</tr>
<tr>
<td>14</td>
<td>UB CF</td>
<td>Isoprime</td>
<td>C2</td>
<td>Y</td>
<td>PCS Drift correction</td>
<td>NBS-19 &amp; IAEA-CO-9 (UCI via INSTAAR)</td>
<td>VSMOW &amp; SLAP (UHEI)</td>
<td>R10, R24, R41, R45</td>
<td>R10, R12, R20, R41</td>
<td>R12, R20, R24, R45</td>
</tr>
<tr>
<td>15</td>
<td>AWI CF</td>
<td>Isoprime</td>
<td>C3</td>
<td>Y</td>
<td>No Drift correction</td>
<td>RM 8562, RM 8563 and RM 8564 (UHEI)</td>
<td>Not measured</td>
<td>R6</td>
<td>R6, R10, R19, R20</td>
<td>Not measured</td>
</tr>
<tr>
<td>16</td>
<td>CIC DI</td>
<td>TC/EA-IRMS &amp; Picarro</td>
<td>(( \delta^{13}C )-CH(_4))</td>
<td>TC/EA-IRMS</td>
<td>( \delta^{15}D )-CH(_4))</td>
<td>Delta V Plus, Delta V Advantage and Picarro</td>
<td>C2</td>
<td>N – – RM 8563</td>
<td>VSMOW &amp; SLAP (UHEI)</td>
<td>R21, R48</td>
</tr>
</tbody>
</table>


| b | Raw ion current correction: the Kr interference was corrected by subtracting the Kr-caused anomalies in the raw ion current. PCS: Kr was separated by a post-combustion separation column or cryogenically. See Sect. 2.5.


2 Overview of atmospheric $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ measurement techniques

2.1 IRMS measurements for $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$

In the 1990s, atmospheric $\delta^{13}$C-CH$_4$ ($\delta$D-CH$_4$) was analysed using an offline technique in which CH$_4$ was separated from the sample air and converted to CO$_2$ (H$_2$) for subsequent offline $\delta^{13}$C-CH$_4$ ($\delta$D-CH$_4$) analyses by dual-inlet isotope ratio mass spectrometry (DI-IRMS; e.g. Stevens and Rust, 1982; Lowe et al., 1991; Quay et al., 1991, 1999; Sugawara et al., 1996; Poß, 2003). The original methodology was based on the combustion of CH$_4$ in sample air, with interfering compounds such as CO$_2$, H$_2$O, N$_2$O, CO and non-methane hydrocarbons being removed cryogenically, chemically or by gas chromatography before CH$_4$ combustion. The number of measurements was limited not only because of laborious and time-consuming laboratory procedures but also because large volumes of air sample were required ($> 100$ LSTP for $\delta$D-CH$_4$). Later, a method based on a continuous-flow gas chromatography isotope ratio mass spectrometry (GC-IRMS) technique combined with combustion and pyrolysis furnaces became available (Merritt et al., 1995; Burgoyne and Hayes, 1998; Hilbert et al., 1999), which dramatically reduced time and effort in the laboratory and likewise the amount of sample air required (now typically $100$ mLSTP).

Such systems are now used in most laboratories worldwide to acquire $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ data in the current and past atmosphere (Rice et al., 2001; Miller et al., 2002; Sowers et al., 2005; Ferretti et al., 2005; Morimoto et al., 2006; Fisher et al., 2006; Behrens et al., 2008; Umezawa et al., 2009; Brass and Röckmann, 2010; Sperlich et al., 2013; Schmitt et al., 2014; Bock et al., 2014; Brand et al., 2016; Röckmann et al., 2016). Although these systems use a similar measurement principle, they vary in the use of pre-concentration of CH$_4$ in sample air, GC separation and combustion/pyrolysis, data corrections and in the specific IRMS instrument among laboratories (see Schmitt et al., 2013, Sect. 3 and Table 1). Besides analysis by mass spectrometry, laser-based spectroscopy has also been developed for atmospheric $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ measurements (Bergamaschi et al., 2000; Eyer et al., 2016), but detailed discussion on the technique is beyond the scope of this study.

2.2 Standard scales

VPDB and VSMOW are the standard scales for $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$, respectively. To make measurements traceable to these standard scales, each laboratory needs to calibrate its laboratory reference gases against reference materials (RMs) with known values on the standard scales. In this study, the term “calibration” means to measure a laboratory gas (for instance a laboratory working standard gas that is routinely compared with samples) against a standard at higher hierarchy level and to assign to that working standard a $\delta^{13}$C-CH$_4$ or $\delta$D-CH$_4$ value traceable to the standard scale. In principle, all measurements at individual laboratories intend to ultimately anchor their working standards and sample gases to the VPDB or VSMOW scale using the RMs provided by the International Atomic Energy Agency (IAEA) or National Institute of Standards and Technology (NIST; Coplen et al., 2006; Brand et al., 2014). However, since RMs and recommended calibration methods for measurements of $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ in air have not yet been provided (Sperlich et al., 2012, 2016), individual groups have developed their own calibration strategies.

Since the $\delta^{13}$C-CH$_4$ measurement by IRMS is taken by $\delta^{13}$C analysis in CO$_2$ oxidized from CH$_4$ in air, some laboratories use pure CO$_2$ gases as a working standard. In many laboratories, these internal CO$_2$ standard gases were calibrated against pure CO$_2$ from the primary anchor of the VPDB-scale NBS-19 or other RMs by using DI-IRMS (Table 1). Since the typical atmospheric $\delta^{13}$C-CH$_4$ value (about $-47\%e$) differs considerably from the $\delta^{13}$C value of NBS-19 ($+1.95\%e$), some laboratories have used other RMs with VPDB values close to atmospheric $\delta^{13}$C-CH$_4$ such as LSVEC (lithium carbonate reference material prepared by Harry J. Svec), IAEA-CO-9 and RM 8563 as a second anchoring point of the VPDB scale (see Table 1). This minimizes the risk of significant errors in realization of the standard scale (due to scale contraction or $^{17}$O correction, described in the following sections). A standard scale established this way at an individual laboratory was often propagated to laboratory-internal CO$_2$ standard gases at lower hierarchy levels, and they were used as the reference in DI-IRMS or GC-IRMS measurement of CO$_2$ processed from CH$_4$ in sample air. Ideally, this accurately links $\delta^{13}$C-CH$_4$ of the sample to the international isotope scale. In contrast, it has been recommended that a measured value of a sample is determined against a reference gas that undergoes the all preparation steps in the sample measurement line in order to cancel out possible isotopic fractionations due to different treatment between the sample and reference gases (principle of identical treatment; Werner and Brand, 2001). This concept has been taken into account in some laboratories; a working standard is calibrated for $\delta^{13}$C-CH$_4$ and sample measurements are referenced by comparison with measurements of that working standard processed in the same manner (e.g. Brand et al., 2016). Despite intentions of best traceability to RMs, the variety of calibrations has resulted in divergent realizations of the VPDB scale across $\delta^{13}$C-CH$_4$ measurement programmes. As in Table 1, the different RMs that have been applied to $\delta^{13}$C-CH$_4$ calibration include NBS-19 (limestone), IAEA-CO-9 (barium carbonate), LSVEC (lithium carbonate) and RM 8562–8564 (CO$_2$); see Coplen et al. (2006), Brand et al. (2014) and Sperlich et al. (2016). It is also noted that uncertainties of assigned values for these RMs range up to a few tenths per mille and the assigned values have been revised over time (Brand et al., 2014), which might have complicated the realization of the standard scale at each labora-
tory. Furthermore, most of these RMs are in different chemical forms, and different isotopic fractionations may have occurred during acid digestion to CO₂, which could have biased calibrations at each laboratory. Lastly, the WMO (2016) has reported exhaustion of NBS-19 and instability of LSVEC, both of which are critical RMs for the VPDB scale. Associated possible revision of δ¹³C values of RMs in the future will affect the consistency of the data sets from different laboratories.

For δD-CH₄, in the conventional offline measurements, CH₄ in sample air needs to be processed to H₂O followed by reduction to H₂ for a subsequent DI-IRMS measurement. GC-IRMS requires pyrolysis of CH₄ to H₂. Therefore, individual laboratories have prepared internal standards of H₂O (liquid) or H₂ (gas), which were calibrated against primary RMs (water) or H₂ reference gases certified for δD (Table 1). Although the situation is less complicated compared to δ¹³C-CH₄ in terms of variety in chemical properties of RMs, the lack of RMs for δD-CH₄ forced laboratories to develop their calibration method standard scale individually. It is also noted that, similarly to δ¹³C-CH₄, this principle of identical treatment has not been followed strictly at all laboratories. If not followed, sample measurements are subject to subtle changes in conditions of the all preparation steps (e.g. conversion of CH₄), while such changes do not affect the measured value of a reference gas injected directly into the IRMS.

2.3 Scale contraction

It has been found that cross contamination between sample and reference CO₂ gases shrinks the δ¹³C distance measured on DI-IRMS (Meijer et al., 2000; Verkouteren et al., 2003a, b). This effect is known as the scale contraction or η effect, and the magnitude is specific to the IRMS instrument and its settings. Since the VPDB scale for δ¹³C-CH₄ has been realized and propagated via CO₂ calibrations by DI-IRMS at individual laboratories, the instrument-dependent scale contraction effect could have caused a significant difference in measurement values, especially at the low δ¹³C values of atmospheric CH₄ of about −47 ‰ (Wendeberg et al., 2013).

2.4 δ¹⁷O correction

For measurement of δ¹³C-CH₄ by IRMS, CH₄ is first oxidized to CO₂ and the different isotopic variants of the produced CO₂ are then registered on Faraday cups with m/z ratio of 44, 45 and 46. Since the raw ion beam intensity for m/z = 45 is the sum of ¹³C¹⁷O₂ and ¹²C¹⁷O¹⁶O, the final δ¹³C value is obtained by correcting for the contribution of the ¹⁷O-containing isotopologue, known as δ¹⁷O correction (e.g. Assonov and Brenninkmeijer, 2003). Several algorithms such as Craig (1957) and Santrock et al. (1985) have been suggested (see Assonov and Brenninkmeijer, 2003 and references therein) and implemented into software/programmes of the IRMS companies and individual laboratories. Assonov and Brenninkmeijer (2003) showed that the bias caused by different ¹⁷O-correction algorithms could exceed general repeatability achieved by IRMS measurements. The ¹⁷O-correction method of each laboratory is listed in Table 1.

2.5 Krypton interference in GC-IRMS

The transition from DI-IRMS to GC-IRMS analyses reduced the analytical effort, but also introduced complications that were initially not recognized and taken into account. It was recently found that atmospheric krypton (Kr) interferes with the δ¹³C-CH₄ GC-IRMS analysis if Kr is present in the ion source during the data acquisition of the CO₂ peak generated from CH₄ oxidation (hereafter CH₄-derived CO₂ peak) (Schmitt et al., 2013). Thus the δ¹³C-CH₄ measurements on a GC-IRMS system can be biased if Kr is not sufficiently separated either from CH₄ or from the CH₄-derived CO₂ peak after the CH₄ combustion. Schmitt et al. (2013) demonstrated that the doubly charged krypton isotope $^{86}_{2+}$Kr, produced in the ion source of an IRMS, can cause lateral tailing extending into the Faraday cups used for δ¹³C analysis (i.e. m/z of 44, 45 and 46), which compromises the measured signal of the CH₄-derived CO₂ peak. This effect had not been recognized for more than a decade since the early years of GC-IRMS measurements (Merritt et al., 1995) and thus has not been taken into account in many data sets of atmospheric δ¹³C-CH₄ reported in the meantime (e.g. Miller et al., 2002; Morimoto et al., 2006; Fisher et al., 2011; Röckmann et al., 2011; Umezawa et al., 2012a, b). Furthermore, because the Kr effect is system dependent and variable with time (Schmitt et al., 2013), applying plausible corrections to past data may not be feasible. Likewise, several gas species including Kr can affect δD-CH₄ measurements, and this effect is also system dependent (Bock et al., 2014).

Several solutions have been suggested to eliminate or account for the Kr interference (Schmitt et al., 2013). Among them, three methods have been implemented at different laboratories (Table 1). Briefly, (1) after the CH₄ oxidation to CO₂, Kr is separated from the CH₄-derived CO₂ by using a post-combustion separation column (PCS) or cryogenically. (2) An offset due to the Kr interference is estimated by comparison with a DI-IRMS measurement (DI offset). (3) The Kr interference peak is subtracted from the raw ion current time series of the IRMS acquisition (raw ion current correction). A more detailed description has been presented in Schmitt et al. (2013).

3 Measurements of participating laboratories

In this section, we briefly document measurement systems of individual laboratories for ease of reference in the following intercomparisons (Sects. 4 and 5). For details, we refer to
more dedicated publications listed in Table 1. The table also visualizes differences among laboratories in terms of possible causes of the measurement offsets described in Sect. 2.

3.1 NIWA

The National Institute for Water and Atmospheric Research (NIWA, originally INS (Institute of Nuclear Sciences) and later INGS (Institute of Nuclear and Geological Sciences)) until 1992) successfully initiated systematic measurements of atmospheric δ13C-CH4 by means of offline CH4 separation and conversion followed by a DI-IRMS measurement in 1988 (Lowe et al., 1988, 1991). A suite of CO2 working gases with δ13C-CH4 values around −47‰ referenced to IAEA materials were utilized to calibrate the measurements. An overall reproducibility of the δ13C-CH4 measurement was evaluated to be 0.02‰ (Lowe et al., 1991). The δ13C-CH4 measurements at NIWA are ultimately calibrated against CO2 produced from NBS-19, IAEA-CO-9 and LSVEC. The long-term δ13C-CH4 records have been presented since then (Lowe et al., 1994, 1997, 2004; Bergamaschi et al., 2001; Schaefer et al., 2016). Bromley et al. (2012) reported that repeated measurements of the two working reference gases and archived air indicated no detectable drift over 16 years since 1992. NIWA has also operated a GC-IRMS system since 2004 (Ferretti et al., 2005) with reproducibility of 0.1‰. The Kr interference on the GC-IRMS δ13C-CH4 measurement has been identified and is corrected by an offset relative to the conventional DI-IRMS measurement (see Sect. 4.1).

3.2 IMAU

The GC-IRMS system at the Institute for Marine and Atmospheric research Utrecht (IMAU) has been described by Brass and Röckmann (2010). The measurement reproducibility is estimated to be 0.07 and 2.3‰ for δ13C-CH4 and δD-CH4, respectively. Sample air is measured against reference air that is processed in the GC-IRMS system in the same manner as a sample. The IMAU δ13C-CH4 standard scale is based on a set of assigned values for 13 firn air samples measured at Max Planck Institute for Chemistry (MPI; Bräunlich et al., 2001) and they are ultimately referenced to a CO2 gas produced from NBS-19 (Röckmann, 1998; Bergamaschi et al., 2000). The δD-CH4 standard scale is based on a set of reference gases originally produced at the MPI (see Sect. 2.3). These calibration details have also been documented by Sperlich et al. (2016). The IMAU system was originally affected by Kr but later modified to remove this interference. A correction was applied for data obtained before the system modification (Schmitt et al., 2013).

3.3 MPIC

The MPIC has reported δ13C-CH4 and δD-CH4 measurements at a baseline station (Bergamaschi et al., 2000) and for firn air samples (Bräunlich et al., 2001) based on an offline DI-IRMS measurement for δ13C-CH4 (Bergamaschi et al., 2000) and a tunable diode laser-based absorption spectrometer (TDLAS) for δD-CH4 (Bergamaschi et al., 1994). Some firn air measurements by Bräunlich et al. (2001) were performed by using a GC-IRMS system at the Laboratory of Glaciology and Geophysics of the Environment. As described in Sect. 3.2, the δ13C-CH4 and δD-CH4 standard scales of MPIC are based on that of IMAU. For the δ13C-CH4 DI-IRMS measurement, the CH4-derived CO2 was measured against a working standard (pure CO2) that was calibrated against NBS-19 on a DI-IRMS system (Röckmann, 1998; Bergamaschi et al., 2000). The MPIC δD-CH4 scale is based on measurements of standard gases at the Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, Germany. CH4 was combusted to CO2 and H2O, followed by reduction of H2O to H2 for subsequent DI-IRMS analysis on H2. They were calibrated against VSMOW and SLAP (Bergamaschi et al., 2000). The measurements of atmospheric δ13C-CH4 and δD-CH4 at the MPIC were discontinued.

3.4 MPI-BGC

The Max Planck Institute for Biogeochemistry (MPI-BGC) set up a GC-IRMS system for δ13C-CH4 and δD-CH4 measurements, and it has been operated for air samples collected at baseline stations (Brand et al., 2016). The long-term (3 years) reproducibility was assessed to be 0.12 for δ13C-CH4 and 1.0‰ for δD-CH4. Initially, the GC-IRMS measurements had been anchored to a working standard air calibrated by IMAU. The Kr effect was eliminated by a PCS column, and the initial calibration has in the meantime been replaced by a new primary calibration, where measurements are ultimately anchored to NBS-19 and LSVEC for δ13C-CH4 and VSMOW-2 and SLAP-2 for δD-CH4 (Sperlich et al., 2016). This calibration, termed JRAS-M16, is the basis for the δ13C-CH4 and δD-CH4 values from MPI-BGC reported in this paper.

3.5 UCI

The University of California Irvine (UCI) measured atmospheric δ13C-CH4 by offline DI-IRMS and δD-CH4 by GC-IRMS (Tyler et al., 1999, 2007; Kai et al., 2011). The UCI GC-IRMS system for both δ13C-CH4 and δD-CH4 has been described in detail by Rice et al. (2001). The measurement reproducibility of the GC-IRMS system was estimated to be 0.05 and 1.5‰ for δ13C-CH4 and δD-CH4, respectively, while that of the offline DI-IRMS δ13C-CH4 measurement was 0.05‰. Samples were measured against laboratory working standard gases of pure CO2 for δ13C-CH4 and pure H2 for δD-CH4. The δ13C-CH4 calibration is based on a CO2 reference gas provided by NIWA, which was compared with CO2 produced from NBS-19 and IAEA-CO-9 (Lowe et al., 1999). The δD-CH4 calibration is referenced to three H2 gas samples measured at a baseline station (Brand et al., 2016). The long-term (3 years) reproducibility was assessed to be 0.12 for δ13C-CH4 and 1.0‰ for δD-CH4. Initially, the GC-IRMS measurements had been anchored to a working standard air calibrated by IMAU. The Kr effect was eliminated by a PCS column, and the initial calibration has in the meantime been replaced by a new primary calibration, where measurements are ultimately anchored to NBS-19 and LSVEC for δ13C-CH4 and VSMOW-2 and SLAP-2 for δD-CH4 (Sperlich et al., 2016). This calibration, termed JRAS-M16, is the basis for the δ13C-CH4 and δD-CH4 values from MPI-BGC reported in this paper.
cylinders purchased from Oztech Gas Company (Rice et al., 2001). The possible Kr interference on the GC-IRMS system is unclear (the laboratory is now closed), but it appears that the Kr effect had been avoided using liquid nitrogen cooling of the GC column as surmised by Schmitt et al. (2013).

3.6 TU

The GC-IRMS system at Tohoku University (TU) has been described by Umezawa et al. (2009). The measurement reproducibility is estimated to be 0.08 for $\delta^{13}$C-CH$_4$ and 2.2‰ for $\delta$D-CH$_4$. Sample measurements are made against pure CO$_2$ and H$_2$ working standard gases for $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$, respectively. The $\delta^{13}$C-CH$_4$ calibration is based on a CO$_2$ primary gas produced from NBS-19. The H$_2$ working standard for the $\delta$D-CH$_4$ measurement is referenced to water laboratory standards that are calibrated against VSMOW and SLAP. Measured $\delta$D-CH$_4$ values are corrected so that the value of a laboratory test gas is kept constant over time to take into account fluctuations in the measured value due to the condition of the pyrolysis furnace (Umezawa et al., 2009, 2012a). The Kr interference in the $\delta^{13}$C-CH$_4$ measurement was identified, but modification or correction has not been implemented. It has been documented that the $\delta^{13}$C-CH$_4$ measurement at TU shifted by $+0.27$‰ after July 2008 (the cause of this sudden shift has yet to be identified) and measurements afterwards were corrected for this value to keep the data consistent (Umezawa et al., 2012a, b). Note that TU made a rigorous re-evaluation of the long-term measurements of their working standard gas recently, and the TU $\delta^{13}$C-CH$_4$ data sets will be revised accordingly. Therefore, the comparison numbers presented here are not comparable to those for earlier publications (Umezawa et al., 2009, 2011, 2012a, b).

3.7 NIPR

The National Institute of Polar Research (NIPR) reported $\delta^{13}$C-CH$_4$ measurements at an Arctic site using a GC-IRMS system (Morimoto et al., 2006, 2017). The measurement reproducibility was evaluated to be 0.06‰. The $\delta^{13}$C-CH$_4$ calibration follows the procedure used by TU. By injecting different quantities of Kr, it was confirmed that ambient Kr does not significantly interfere with the $\delta^{13}$C-CH$_4$ measurements at NIPR.

3.8 UW

The University of Washington (UW) reported extensive $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ measurements using an offline DI-IRMS system (Quay et al., 1991, 1999). The reproducibility was estimated to be 0.1 for $\delta^{13}$C-CH$_4$ and 3–4‰ for $\delta$D-CH$_4$. The $\delta^{13}$C-CH$_4$ calibration is based on measurements made against NBS-19 (Quay et al., 1999), while the earlier measurements were calibrated against NBS-20 and NBS-16 (Quay et al., 1991). The $\delta$D-CH$_4$ was anchored to calibration by VSMOW and SLAP. Systematic measurements of air standards showed that no significant time shift ($+0.001 \pm 0.002$‰ yr$^{-1}$) affected their $\delta^{13}$C-CH$_4$ data set for 1988–1995 (Quay et al., 1999).

3.9 UHEI

The University of Heidelberg (UHEI) carried out $\delta^{13}$C-CH$_4$ measurements by DI-IRMS (Levin et al., 1999, 2012). The typical measurement reproducibility was evaluated to be 0.05‰ (Levin et al., 1999). The UHEI $\delta^{13}$C-CH$_4$ measurements are calibrated against CO$_2$ reference materials (RM 8562, RM 8563 and RM 8564; Behrens et al., 2008). Although reported previously only for signatures of source CH$_4$ (Levin et al., 1993), UHEI also took offline $\delta$D-CH$_4$ measurements of atmospheric samples by DI-IRMS and TDLAS (Poß, 2003). The $\delta$D-CH$_4$ measurements by DI-IRMS were taken for pure H$_2$ (H$_2$O from CH$_4$ oxidation converted to H$_2$ with zinc as catalyst) and were calibrated against VSMOW and SLAP. Note that UHEI recently re-evaluated all their atmospheric $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ measurements rigorously, based on the history of laboratory standards used; therefore, comparison numbers published in earlier works are not comparable to the revised values presented here.

3.10 INSTAAR

The Institute of Arctic and Alpine Research (INSTAAR) of the University of Colorado, Boulder has measured $\delta^{13}$C-CH$_4$ and, intermittently, $\delta$D-CH$_4$ using a GC-IRMS system for flask air samples from the cooperative sampling network of National Oceanic and Atmospheric Administration (NOAA; Miller et al., 2002). Reproducibilities of the $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ measurements are evaluated to be 0.08 and 2‰, respectively (Miller et al., 2002; White et al., 2016). The INSTAAR $\delta^{13}$C-CH$_4$ measurement currently follows the UCI calibration, while the $\delta$D-CH$_4$ measurement is not explicitly anchored to the VSMOW scale (White et al., 2016). The Kr interference in the $\delta^{13}$C-CH$_4$ measurement is significant, and a PCS column was therefore implemented into the system in May 2017. Correction of the data for the Kr interference (1998–present) is under evaluation. Of the data presented here, only the ice core intercomparison round robin (Sect. 3.4) and the INSTAAR–MPI-BGC comparison (Sect. 3.5) have not been interfered with by Kr.

3.11 RHUL

Royal Holloway University of London (RHUL) measured atmospheric $\delta^{13}$C-CH$_4$ using an offline DI-IRMS technique (Lowry et al., 2001) and a GC-IRMS system (Fisher et al., 2006, 2011; Nisbet et al., 2016). Reproducibility of the DI-IRMS measurement was evaluated to be 0.04‰ (Lowry et al., 2001) and that by the GC-IRMS is 0.05‰ (Fisher et al., 2006). They ultimately calibrated $\delta^{13}$C-CH$_4$ to IAEA carbonate materials NBS-19 and IAEA-CO-9 (Lowry et al.,
Portland State University (PDX) reported δ^{13}C-CH_4 and δD-CH_4 measurements for archive air samples (Rice et al., 2016). The PDX measurement system has been described in Teama (2013) with some updates since Rice et al. (2001). The δ^{13}C-CH_4 and δD-CH_4 reproducibilities are 0.07 and 2.0‰, respectively, and PDX shares the standard scales with UCI for both δ^{13}C-CH_4 and δD-CH_4 (Rice et al., 2016).

3.13 PSU

Pennsylvania State University (PSU) reported δ^{13}C-CH_4 and δD-CH_4 data from ice cores and firn air using a GC-IRMS system (e.g. Sowers et al., 2005; Sowers, 2010). The overall measurement reproducibility including every step for ice core measurements was evaluated to be 0.3 for δ^{13}C-CH_4 and 3‰ for δD-CH_4 (Sowers, 2010). The PSU δ^{13}C-CH_4 measurements are calibrated against CO_2 RMs (RM 8563 and RM 8564). The δD-CH_4 calibration is against H_2 gas bottles from Oztech Gas Company (Sowers, 2006).

3.14 UB

The University of Bern (UB) makes δ^{13}C-CH_4 measurements from ice cores using a GC-IRMS system with an overall reproducibility of 0.15‰ (Schmitt et al., 2014; Bock et al., 2017). The UB measurements are referenced to a whole-air working standard with a CH_4 mole fraction of 1508.2 ppb and an assigned δ^{13}C-CH_4 value of −47.34 ± 0.02‰ (named “Boulder, CA08289” in Schmitt et al., 2014). This value is anchored to the standard scale used at INSTAAR (Sect. 3.10). UB also measures δD-CH_4 for ice core samples (Bock et al., 2010, 2014, 2017). The overall measurement precision for ice core sample (including extraction of air from an ice sample) was evaluated to be 2.3‰. The UB δD-CH_4 measurement is referenced by using an ambient air cylinder (named “Air Control©”) with a δD-CH_4 value of −93.6 ± 2.8‰, which was cross-referenced to a high-pressure cylinder filled at the Alert Station (“Alert 2002/11” with δD-CH_4 of −82.2 ± 1.0‰) analysed on the scale maintained at UHEI (Bock et al., 2010, 2014). However, this value has to be corrected to −85.2 ± 1.0‰ to account for the recent re-evaluation at UHEI (Sect. 3.9). All UB data published after 2011 are free of Kr interference.

3.15 AWI

The Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research (AWI) reported δ^{13}C-CH_4 measurements from ice cores using a GC-IRMS system (Behrens et al., 2008; Fischer et al., 2008; Möller et al., 2013). The measurement reproducibility was estimated to be 0.2‰. The δ^{13}C-CH_4 measurements employed the UHEI standard scale via comparison of measurements of an air sample collected at Neumayer Station, Antarctica (Möller et al., 2013).

3.16 CIC

TheCentre for Ice and Climate (CIC) of the Niels Bohr Institute has reported δ^{13}C-CH_4 measurements from ice cores (Sperlich et al., 2015) using a GC-IRMS system with measurement reproducibility of 0.09‰ (Sperlich et al., 2013). CIC also set up an offline combustion system for samples with a large amount of CH_4, which is combined with DI-IRMS for δ^{13}C-CH_4 and with either a high temperature conversion/elemental analyser (TC/EA) coupled to IRMS or laser spectroscopy for δD-CH_4 (Sperlich et al., 2012); the measurement reproducibility is 0.04 for δ^{13}C-CH_4 and 0.7‰ for δD-CH_4. The CIC measurements are referenced to RM 8563 for δ^{13}C-CH_4 and VSMOW-2 and SLAP-2 for δD-CH_4. The combined uncertainty of this analytical system including the uncertainty of the entire traceability chain was estimated at 0.07 for δ^{13}C-CH_4 and 0.7‰ for δD-CH_4 (Sperlich et al., 2016).

4 Intercomparison exercises

4.1 Intercomparison between UCI and IMAU

An intercomparison between UCI and IMAU was made by analysing six air samples at both laboratories; the air samples were collected along a flight track of commercial aircraft in the upper troposphere in the early phase of the CARIBIC (Civil Aircraft for the Regular Investigation of the atmosphere Based on an Instrument Container) project (Brenninkmeijer et al., 1999). The original samples were collected into large stainless steel cylinders (21 L) and aliquots of them were transferred into smaller stainless steel canisters (∼2.3 L) for storage after delivery to the MPIC laboratory. Different subsamples from identical original samples were sent to UCI and IMAU for analysis, and they were measured at UCI in 2008 and at IMAU in 2012 to 2013. The measurement results at both laboratories are summarized in Table 2. The result indicated significant differences of +0.42 ± 0.04 ‰ for δ^{13}C-CH_4 (UCI value is higher than at IMAU) and of −10.7 ± 0.7 ‰ for δD-CH_4 (UCI value is lower than IMAU).

4.2 Intercomparison between TU/NIPR and IMAU

An intercomparison between TU/NIPR and IMAU was carried out during 2013–2015. The TU laboratory prepared four stainless steel canisters (∼1 L) filled with dried ambient air (canisters MD1 and MD2) and CH_4-in-synthetic air gas (canisters MD3 and MD4) with CH_4 mole fractions ranging from 899 to 2117 ppb on the TU CH_4 scale (Aoki et al., 1992;
Table 2. Result of intercomparison of $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ measurements between UCI and IMAU.

| Sample ID | CH$_4$ (ppb)$^a$ | $\delta^{13}$C-CH$_4$ (%)$^b$ | | | $\delta$D-CH$_4$ (%)$^b$ |
|-----------|-----------------|------------------------------|-----------------|-----------------|
|           | UCI             | IMAU | UCI–IMAU | UCI             | IMAU | UCI–IMAU |
| WAS-24-2  | 1784.7          | $-46.96 \pm 0.07$ (N = 3) | $-47.33 \pm 0.05$ (N = 3) | +0.37 | $-91.6 (1.0, N = 2)$ | $-78.9 \pm 0.1 (N = 4)$ | $-12.7$ |
| WAS-24-5  | 1825.8          | $-47.16$ (N = 1) | $-47.53 \pm 0.02$ (N = 6) | +0.37 | $-93.8 (N = 1)$ | $-83.1 \pm 0.2 (N = 4)$ | $-10.7$ |
| WAS-24-6  | 1827.5          | $-47.08$ (0.02, N = 2) | $-47.55 \pm 0.04$ (N = 6) | +0.47 | $-92.1 (1.6, N = 2)$ | $-83.6 \pm 0.1 (N = 4)$ | $-8.5$ |
| WAS-24-9  | 1799.8          | $-47.05$ (N = 1) | $-47.38 \pm 0.02$ (N = 6) | +0.33 | $-92.3 \pm 1.8 (N = 3)$ | $-79.8 \pm 0.8 (N = 4)$ | $-12.4$ |
| WAS-24-10 | 1789.8          | $-47.07$ (N = 1) | $-47.42 \pm 0.02$ (N = 6) | +0.35 | $-89.3 (N = 1)$ | $-79.7 \pm 0.8 (N = 4)$ | $-9.6$ |
| WAS-24-11 | 1780.8          | $-46.77$ (N = 1) | $-47.37 \pm 0.03$ (N = 6) | +0.60 | $-89.0 (1.8, N = 2)$ | $-78.7 \pm 0.7 (N = 4)$ | $-10.3$ |
| Average   |                 | $+0.42 \pm 0.04$ | | | $-10.7 \pm 0.7$ |

$^a$ NOAA-2004 CH$_4$ scale (Dlugokencky et al., 2005). $^b$ Uncertainties are standard errors of the mean for measurements with N $\geq$ 1. Difference in duplicate flask measurements (N = 2) is shown in parenthesis.

Table 3. Result of intercomparison of $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ measurements between TU/NIPR and IMAU.

| Sample ID | CH$_4$ (ppb)$^a$ | $\delta^{13}$C-CH$_4$ (%)$^b$ | | | $\delta$D-CH$_4$ (%)$^b$ |
|-----------|-----------------|------------------------------|-----------------|-----------------|
|           | TU              | NIPR | IMAU | Difference from | TU              | NIPR | IMAU | Difference from |
|           |                 |      |      | IMAU           |                 |      |      | IMAU           |
| MD1       | 1901.1          | $-47.04 \pm 0.02$ (N = 16) | $-47.11 \pm 0.02$ (N = 5) | $-47.40 \pm 0.04$ (N = 9) | +0.36 (TU) | $-97.2 \pm 0.6$ (N = 10) | $-85.0 \pm 0.1$ (N = 8) | $-12.2$ |
| MD2       | 2116.6          | $-46.81 \pm 0.02$ (N = 16) | $-46.92 \pm 0.03$ (N = 6) | $-47.26 \pm 0.03$ (N = 9) | +0.45 (TU) | $-118.5 \pm 0.6$ (N = 10) | $-104.5 \pm 0.3$ (N = 8) | $-14.0$ |
| MD3       | 899.1           | $-41.14 \pm 0.04$ (N = 16) | $-41.05 \pm 0.02$ (N = 6) | $-41.81 \pm 0.03$ (N = 9) | +0.34 (NIPR) | $-190.7 \pm 0.6$ (N = 10) | $-175.8 \pm 0.6$ (N = 8) | $-14.9$ |
| MD4       | 1700.5          | $-42.47 \pm 0.03$ (N = 16) | $-42.43 \pm 0.04$ (N = 6) | $-42.98 \pm 0.02$ (N = 9) | $+0.76$ (NIPR) | $-195.2 \pm 0.6$ (N = 10) | $-180.6 \pm 0.2$ (N = 8) | $-14.6$ |
| Average   |                 | $+0.40 \pm 0.04$ (TU)$^c$ | $+0.51 \pm 0.03$ (NIPR)$^c$ | | $-13.1 \pm 0.6$ |

$^a$ Tokyo University CH4 scale (Aoki et al., 1992; Umezawa et al., 2014). $^b$ Uncertainties are standard errors of the mean for the repetitive measurements. $^c$ Uncertainties are standard errors of the mean for differences in the above lines.

Umezawa et al., 2014; Table 3). The canisters were analysed at TU and then sent to IMAU, after which they were sent back to TU and reanalysed to confirm the stability of the air samples in the canisters during the intercomparison exercise. The measurements at TU before and after the transport to IMAU from April 2013 to July 2015 indicated that possible drifts during canister storage and transportation are small (<0.1 for $\delta^{13}$C-CH$_4$ and <3.5‰ for $\delta$D-CH$_4$). NIPR also measured the canisters for $\delta^{13}$C-CH$_4$. The results indicate significant differences of $+0.50 \pm 0.07$‰ for $\delta^{13}$C-CH$_4$ (TU value is higher than IMAU) and of $-13.9 \pm 0.9$‰ for $\delta$D-CH$_4$ (TU value is lower than IMAU; Table 3). The measurements of the four canisters at NIPR were $+0.48 \pm 0.11$‰ higher than IMAU. However, the differences of $\delta^{13}$C-CH$_4$ measurements are smaller for the ambient air samples (MD1 and MD2) than the CH$_4$-in-synthetic air samples (MD3 and MD4). It is also noted that the $\delta^{13}$C-CH$_4$ difference between the laboratories is largest for the low CH$_4$ mole fraction (~900 ppb) sample (MD3). The cause is unclear, but might be related to (1) deviation in $\delta^{13}$C-CH$_4$ of the latter samples from the typical atmospheric value, i.e. scale contraction effect; (2) difference in air matrix, i.e. natural versus synthetic air; and (3) difference in linearity with respect to CH$_4$ mole fraction. This result therefore indicates that the measurement offset is not constant for a wide range of $\delta^{13}$C-CH$_4$ values and CH$_4$ mole fractions or for differences in the air matrix. Since we focus in this study on comparison of atmospheric samples, the intercomparison results for the ambient air samples are considered as interlaboratory measurement offsets. The average differences for ambient air are $+0.40 \pm 0.04$ for TU and $+0.31 \pm 0.03$ for NIPR relative to IMAU. Likewise, the $\delta$D-CH$_4$ offset of TU versus IMAU is considered to be $-13.1 \pm 0.6$‰.

4.3 Intercomparison between UHEI and MPI-BGC

An intercomparison between UHEI and MPI-BGC was conducted in 2013 on six archived air samples from Neumayer station, Antarctica. These samples, collected in the time period from 1988 to 2008, had been analysed by UHEI for $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ by D-IRMS (two samples were analysed for $\delta$D-CH$_4$ additionally by TDLAS) during 2003–2010 and were stored in high-pressure cylinders. The typical reproducibility for the measurements is between 0.02 and 0.05‰ for $\delta^{13}$C-CH$_4$ and between 1.6 and 2.6‰ for $\delta$D-CH$_4$. In 2013, duplicate aliquots were sampled in 1 L glass
Table 4. Result of intercomparison of $\delta^{13}$C and $\delta$D-CH$_4$ measurements between UHEI and MPI-BGC. n.a.: not analysed.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Preparation Date</th>
<th>Analysis Date</th>
<th>UHEI  $\delta^{13}$C-CH$_4$ (‰)</th>
<th>MPI-BGC  $\delta^{13}$C-CH$_4$ (‰)</th>
<th>UHEI–MPI-BGC $\delta^{13}$C-CH$_4$ (‰)</th>
<th>UHEI  $\delta$D-CH$_4$ (%)</th>
<th>MPI-BGC  $\delta$D-CH$_4$ (%)</th>
<th>UHEI–MPI-BGC $\delta$D-CH$_4$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GvN 88/20</td>
<td>(24 Jul 1988)</td>
<td>17 Dec 2003</td>
<td>17 Sep 2013</td>
<td>$-47.54$</td>
<td>$-47.66$</td>
<td>$+0.13$</td>
<td>$-83.3$</td>
<td>$-82.1$</td>
</tr>
<tr>
<td>GvN 92/12</td>
<td>(11 May 1992)</td>
<td>11 Dec 2008</td>
<td>17 Jun 2013</td>
<td>$-47.43$</td>
<td>$-47.40$</td>
<td>$-0.03$</td>
<td>$-79.1$</td>
<td>$-81.2$</td>
</tr>
<tr>
<td>GvN 96/03</td>
<td>(13 Feb 1996)</td>
<td>11 Nov 2003</td>
<td>17 Jun 2013</td>
<td>$-47.27$</td>
<td>$-47.18$</td>
<td>$-0.09$</td>
<td>$-73.9$</td>
<td>$-74.6$</td>
</tr>
<tr>
<td>GvN 99/14</td>
<td>(29 Dec 1999)</td>
<td>3 Apr 2003</td>
<td>9 Jul 2013</td>
<td>$-47.30$</td>
<td>$-47.23$</td>
<td>$-0.07$</td>
<td>$-75.2$</td>
<td>$-74.6$</td>
</tr>
<tr>
<td>GvN 06/14</td>
<td>(23 Sep 2006)</td>
<td>7 May 2003</td>
<td>9 Jul 2013</td>
<td>n.a.</td>
<td>$-47.19$</td>
<td>$-0.09$</td>
<td>$-72.3$</td>
<td>$-73.1$</td>
</tr>
<tr>
<td>GvN 08/03</td>
<td>(6 Mar 2008)</td>
<td>28 Jul 2010</td>
<td>17 Jun 2013</td>
<td>$-47.18$</td>
<td>$-47.35$</td>
<td>$+0.17$</td>
<td>n.a.</td>
<td>$-67.4$</td>
</tr>
</tbody>
</table>

Average $+0.02 \pm 0.05^b$ $+0.4 \pm 0.6^b$

$^a$ Difference in duplicate flask measurements is shown in parenthesis. $^b$ Uncertainties are standard errors of the mean for differences in the above lines.

flasks and analysed at MPI-BGC. The measurement results at both laboratories are summarized in Table 4. The results show insignificant measurement offsets of $+0.02 \pm 0.05$ for $\delta^{13}$C-CH$_4$ and of $+0.4 \pm 0.6$‰ for $\delta$D-CH$_4$ (with the MPI-BGC values being more negative than those from UHEI in both cases).

4.4 Round robin comparison of ice core analysis laboratories

A round robin cylinder exercise was initiated to facilitate intercomparison of laboratories that measure $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ in ice core and firm air samples. Part of this exercise has been presented previously (Table 2 in Schmitt et al., 2013). Three high-pressure Al cylinders were filled with varying trace gas compositions to mimic present day, pre-industrial and last-glacial air mole fractions. The CH$_4$ mole fractions of these cylinders were 1830.6 (CA 03560), 904.0 (CC 71560) and 372.2 ppb (CA 01179) on the NOAA-2004 CH$_4$ scale (Dlugokencky et al., 2005), respectively. The cylinders were shipped to the laboratories listed in Table 5 for analysis of all constituents that each lab was capable of measuring. Part of this exercise has been presented previously (Table 2 in Schmitt et al., 2013). In some cases, aliquots from the tanks were measured using offline combustion to CO$_2$ followed by $\delta^{13}$C-CH$_4$ analyses via conventional DI-IRMS. The cylinders were re-measured at PSU at the end of the round robin to verify that the isotopic composition had not shifted over the 9 years during the transportation of the cylinders. The difference between the 2007 and 2016 $\delta^{13}$C-CH$_4$ measured at PSU was less than 0.14‰ for two of the three cylinders, indicating that the isotopic composition of the cylinder air was stable throughout the intercomparison exercise. The third cylinder (CA 01179) was 0.58‰ off from the original measurement, which is just outside the analytical uncertainty associated with PSU measurements. There may have been a slight drift over the 9 years between measurements, although the cause has yet to be resolved. The results of the $\delta^{13}$C-CH$_4$ intercomparison showed agreement with the average standard deviation among all six participating laboratories better than 0.37‰ for the cylinders with high (CA 03560) and middle (CC 71560) mole fractions. $\delta$D-CH$_4$ results show more scatter due to the difficult nature of the measurements and the offset among the standard scales.

4.5 Intercomparison between INSTAAR and MPI-BGC

An intercomparison between INSTAAR and MPI-BGC was recently made by analysing three air cylinders at both laboratories. They were measured at MPI-BGC between April and July of 2016 and at INSTAAR between May and June of 2017. Two of the cylinders have ambient CH$_4$ mole fractions (3000 ppb; HUEY-001 and DEWY-001) and the other has a lower value (1500 ppb; LOUI-001; Table 6). In addition, air from another suite of cylinders was sampled into flasks at INSTAAR and sent to MPI-BGC. Measurements at MPI-BGC and INSTAAR were taken in January–February of 2017 and May–June of 2017, respectively. The four cylinders (CART-001, STAN-001, KENN-001 and KYLE-001) have different CH$_4$ mole fractions and $\delta^{13}$C-CH$_4$ values. The measurement results are summarized in Table 6. The INSTAAR data presented here were not interfered with by Kr by installing a PC column into the system. The results show significant but consistent measurement offsets of $+0.28 \pm 0.01$‰ for the five cylinders with different CH$_4$ mole fractions and ambient $\delta^{13}$C-CH$_4$ values (with the IN-
STAAR values being more positive than those from MPI-BGC. The measurements for the cylinder with low $\delta^{13}$C-CH$_4$ values were 0.60‰ ± 0.9 between both laboratories presumably due to the scale contraction effect. It is noted that the INSTAAR measurements without the Kr removal yielded a higher $\delta^{13}$C-CH$_4$ value (+0.44 ± 0.02‰ relative to the MPI-BGC measurement) for one cylinder (LOUI-001), which presumably reflects the Kr interference pronounced at a lower CH$_4$ mole fraction.

4.6 Intercomparison based on co-located samples through the NOAA cooperative sampling network

The Cooperative Flask Sampling Network, operated by the NOAA Global Monitoring Division, collects air samples from numerous sites around the world, and INSTAAR has analysed those air samples for $\delta^{13}$C-CH$_4$ since 1998. There are several sites at which air samples have been concurrently collected by other laboratories. RHUL has analysed air samples at Alert (ALT), Canada and Ascension Island (ASC), and NIWA has done at Baring Head (BHD), New Zealand. Although the individual laboratories do not measure the same sample air in these cases, these co-located air samples provide an opportunity for assessment of possible measurement uncertainties.

### Table 5. Results from the round robin comparison of $\delta^{13}$C and $\delta^{15}$D measurements of CH$_4$

<table>
<thead>
<tr>
<th>Laboratory</th>
<th>CA 03560 (1830.6 ppb)</th>
<th>CA 71560 (904.0 ppb)</th>
<th>CA 01179 (372.2 ppb)</th>
<th>Kr corr.</th>
<th>Analysis Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSU</td>
<td>$-47.20 \pm 0.16$</td>
<td>$-47.41 \pm 0.10$</td>
<td>$-47.52 \pm 0.06$</td>
<td>Raw ion current correction$^a$</td>
<td>Jul 2007</td>
</tr>
<tr>
<td>UCI</td>
<td>$-46.96 \pm 0.16$</td>
<td>$-47.20 \pm 0.10$</td>
<td>$-47.41 \pm 0.12$</td>
<td>DI offset$^b$</td>
<td>Jul 2007</td>
</tr>
<tr>
<td>INSTAAR (DI-IRMS)</td>
<td>$-47.10 \pm 0.05$</td>
<td>$-47.09 \pm 0.06$</td>
<td>$-46.83 \pm 0.12$</td>
<td>PCS$^c$</td>
<td>May 2016</td>
</tr>
<tr>
<td>NIWA</td>
<td>$-47.09 \pm 0.12$</td>
<td>$-47.40 \pm 0.08$</td>
<td>$-47.23 \pm 0.06$</td>
<td>–</td>
<td>Dec 2007$^e$</td>
</tr>
<tr>
<td>NIMAU</td>
<td>$-47.08 \pm 0.05$</td>
<td>$-47.20 \pm 0.06$</td>
<td>$-46.78 \pm 0.06$</td>
<td>PCS$^c$</td>
<td>Dec 2008</td>
</tr>
<tr>
<td>IMAU</td>
<td>$-47.23 \pm 0.02$</td>
<td>$-47.44 \pm 0.02$</td>
<td>$-47.43 \pm 0.02$</td>
<td>–</td>
<td>Jun 2009</td>
</tr>
<tr>
<td>NIWA (GC-IRMS)</td>
<td>$-47.44 \pm 0.21$</td>
<td>$-48.34 \pm 0.28$</td>
<td>$-47.62 \pm 0.11$</td>
<td>DI offset$^b$</td>
<td>Jun 2009</td>
</tr>
<tr>
<td>UB</td>
<td>$-47.41 \pm 0.09$</td>
<td>$-47.37 \pm 0.07$</td>
<td>$-47.31 \pm 0.11$</td>
<td>No$^d$</td>
<td>Jan 2011</td>
</tr>
<tr>
<td></td>
<td>± 80.4 ± 2.2</td>
<td>± 81.0 ± 2.0</td>
<td>± 86.2 ± 3.3</td>
<td>± 78.8 ± 12.4</td>
<td>Dec 2010–Jan 2011</td>
</tr>
</tbody>
</table>

$^a$ Raw ion current correction: the Kr interference was corrected by subtracting the Kr-caused anomalies in the raw ion current data (Sect. 5.4 of Schmitt et al., 2013). $^b$ DI offset: the Kr interference was corrected by an offset relative to a DI-IRMS measurement. $^c$ PCS: Kr was separated by a post-combustion separation column (Sect. 5.2 of Schmitt et al., 2013). $^d$ No: Measurements are affected by the Kr interference (old system without PCS) and raw ion current correction was not possible. $^e$ Estimated because no exact record on the analysis date at UCI is unfortunately available. $^f$ Uncertainties are standard deviations of multiple measurements at each laboratory.

### Table 6. Results of a $\delta^{13}$C-CH$_4$ intercomparison between INSTAAR and MPI-BGC.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>CH$_4$ (ppb)$^a$</th>
<th>$\delta^{13}$C-CH$_4$ (%)$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>INSTAAR</td>
</tr>
<tr>
<td>HUEY-001</td>
<td>1905.5</td>
<td>$-47.37 \pm 0.01$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(N = 22)</td>
</tr>
<tr>
<td>DEWY-001</td>
<td>1879.9</td>
<td>$-47.38 \pm 0.01$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(N = 26)</td>
</tr>
<tr>
<td>LOUI-001</td>
<td>1496.0</td>
<td>$-47.26 \pm 0.01$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(N = 17)</td>
</tr>
<tr>
<td>CART-001</td>
<td>1848.1</td>
<td>$-42.98 \pm 0.01$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(N = 21)</td>
</tr>
<tr>
<td>STAN-001</td>
<td>1696.4</td>
<td>$-56.60 \pm 0.01$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(N = 7)</td>
</tr>
<tr>
<td>KENN-001</td>
<td>1847.6</td>
<td>$-47.65 \pm 0.01$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(N = 26)</td>
</tr>
<tr>
<td>KYLE-001</td>
<td>1847.6</td>
<td>$-47.27 \pm 0.01$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(N = 29)</td>
</tr>
</tbody>
</table>

$^a$ NOAA-2004 CH$_4$ scale (Dlugokencky et al., 2005). $^b$ Uncertainties are standard errors of the mean for the repetitive measurements.
offssets as examined previously (Levin et al., 2012). (1) For the RHUL–INSTAAR difference, the $\delta^{13}$C-CH$_4$ data at ALT during 2009–2014 and at ASC during 2010–2015 were compared to each other if both air samples were collected within a 10 h interval. The ALT and ASC comparisons indicated that the INSTAAR measurement is $+0.05 \pm 0.16 (N = 350)$ and $0.00 \pm 0.17 \%e (N = 80)$ higher than RHUL, respectively. Note that, for this comparison, the RHUL GC-IRMS data were corrected by $-0.20 \%e$; the offset value was estimated from measurements of flasks filled from two different cylinders (CH$_4$ in air, both at ambient mole fraction level, one at ambient $\delta^{13}$C-CH$_4$ and the other at about $-56 \%e$ by spiking $^{13}$C-depleted CH$_4$). (2) For the NIWA–INSTAAR comparison, the $\delta^{13}$C-CH$_4$ data at BHD during 2009–2014 from both laboratories were compared if both air samples were collected within a 15 h interval. The result indicates that the INSTAAR measurement is $+0.08 \pm 0.11 \%e (N = 45)$ higher than NIWA.

5 Measurement offsets among laboratories

Here we revisit intercomparisons published previously. Some laboratories employed a standard scale from another laboratory. Such intercomparisons and interlaboratory scale propagations reported in the literature are displayed in Fig. 1. In this section we review the previous and present intercomparisons and accordingly suggest plausible measurement offsets among different laboratories (Fig. 2). Relevant information is summarized in Table 1 and the subsections below correspond to those in Sect. 3. Since some laboratories focus on $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ measurements from ice core and firn air samples to elucidate changes of atmospheric CH$_4$ in the past, Fig. 2 also combines $\delta^{13}$C-CH$_4$ and $\delta$D-CH$_4$ data both for the modern and past atmosphere. It is, however, noted that Fig. 2 suggests the measurement offsets at the modern CH$_4$ mole fraction and isotopic ratios that such values could be different for the past atmosphere (see Sect. 4.2, 4.4 and 4.5).

In this study, we report $\delta^{13}$C-CH$_4$ offsets with respect to the conventional DI-IRMS measurement at NIWA (Lowe et al., 1991) because NIWA’s $\delta^{13}$C-CH$_4$ measurements have been compared with those from the most laboratories to date (Table 1 and Fig. 1). In contrast, $\delta$D-CH$_4$ measurements from different laboratories have been limited. We report $\delta$D-CH$_4$ offsets of different laboratories with respect to the IMAU measurement. The uncertainties presented in this study are generally standard errors of the mean, but numbers in the literature are cited as is. It should be therefore noted that the uncertainties, in particular those calculated by error propagation, are not rigorously consistent in all places in the paper.

5.1 NIWA

5.1.1 $\delta^{13}$C-CH$_4$

As listed in Table 1, the DI-IRMS measurement at NIWA has been repeatedly intercompared with other laboratories. Importantly for this comparison, Bromley et al. (2012) reported the long-term stability of the measurement over the years 1992–2007, and it is likewise confirmed until 2011. The NIWA GC-IRMS system, based on the methodology of Miller et al. (2002), has an offset relative to the DI-IRMS of $-0.19 \pm 0.26 \%e$. Measurements on the GC-IRMS informing this instrument comparison are subject to the Kr interference. A Kr-correction has since been derived in an empirical equation from the round robin intercomparison results (Schmitt et al., 2013 and Sect. 4.4), accounting for differences in the CH$_4$ mole fraction and an exponential fit to the GC-IRMS versus DI-IRMS results. The GC-IRMS system is currently equipped with a PCS column to eliminate the Kr interference.

5.2 IMAU

5.2.1 $\delta^{13}$C-CH$_4$

According to Schmitt et al. (2013), the IMAU measurement at the present CH$_4$ mole fraction level is in agreement with NIWA with an offset value of $-0.04 \pm 0.07 \%e$ (no. 2 in Fig. 2a). This corresponds to the round robin comparison for the cylinder with a CH$_4$ mole fraction of 1830.6 ppb (CA 03560) in Table 5 (Sect. 4.4). The difference is $-0.03 \pm 0.05 \%e$ for data analysed before the modification to remove the Kr interference (see Table 2 in Schmitt et al., 2013). The intercomparison in this study (Sect. 3.4) also shows that the IMAU offset is $-0.08 \pm 0.11 \%e$ for the cylinder with the CH$_4$ mole fraction of 904.0 ppb (CA 71560).

5.2.2 $\delta$D-CH$_4$

As listed in Table 1, IMAU has made the most intercomparisons with other laboratories so far. It is noted that the standard scale at IMAU was propagated from the MPIC (Bergamaschi et al., 2000; Sect. 2.2) and that it recently showed a reasonable agreement with the recent calibration at MPI-BGC (Sperlich et al., 2016).

5.3 MPIC

5.3.1 $\delta^{13}$C-CH$_4$

As written in Sect. 3.3, the standard scale at the MPIC was transferred to IMAU (Brass and Röckmann, 2010; Sperlich et al., 2016). Since no direct comparison with NIWA is available, the MPIC offset relative to NIWA is estimated to be $-0.04 \pm 0.07 \%e$, identical to the IMAU offset (no. 3 in Fig. 2a).
5.3.2 \( \delta^{13}\text{C-CH}_4 \)

Bock et al. (2010) reported an intercomparison using firn air samples between the UB and MPIC, which indicated that, combined with the UB \( \delta^{13}\text{C-CH}_4 \) offset (Sect. 5.14), the MPIC \( \delta^{13}\text{C-CH}_4 \) offset is \(+0.3 \pm 1.1\ \% e\) with respect to IMAU (no. 3 in Fig. 2b).

5.4 MPI-BGC

5.4.1 \( \delta^{13}\text{C-CH}_4 \)

Sperlich et al. (2016) quantified the offset of the IMAU standard scale relative to the primary standard scale at MPI-BGC. It was indicated that the MPI-BGC measurement differs by \(-0.03 \pm 0.10\ \% e\) from the IMAU standard scale. Combined
with the IMAU offset relative to NIWA (Sect. 5.2), the MPI-BGC offset is estimated to be \(-0.07 \pm 0.12\%e\) (no. 4 in Fig. 2a).

5.4.2 \(\delta^13\text{C-Ch}_4\)

According to Sperlich et al. (2016), the MPI-BGC measurement is \(-4.2 \pm 1.2\%e\) from IMAU (no. 4 in Fig. 2b).

5.5 UCI

5.5.1 \(\delta^13\text{C-Ch}_4\)

Intercomparison exercises of UCI with external laboratories have been made several times. The oldest intercomparison (Lowe et al., 1991) reported good agreement (<0.02\%e) between the former UCI laboratory (S. Tyler at NCAR) and NIWA (INS, IGNS at that time). Among the later measurements, there were two direct intercomparisons with NIWA. (1) Tyler et al. (2007) reported an intercomparison result of UCI to be \(-0.01 \pm 0.09\%e\) with respect to NIWA (top in no. 5, Fig. 2a). For this comparison, 16 air samples collected at Niwot Ridge, Colorado or Baring Head, New Zealand were exchanged between UCI and NIWA in 1998–1999. (2) This study (Sect. 4.4 and Table 5) shows that the UCI measurement are \(+0.14 \pm 0.12\%e\) (bottom of no. 5 in Fig. 2a) and \(+0.04 \pm 0.08\%e\) higher than NIWA for the cylinders with high (CA 03560) and middle (CC 71560) \text{CH}_4 mole fractions, respectively. (3) In contrast, the intercomparison in this study (Sect. 4.1 and Table 2) combined with the IMAU offset (Sect. 5.2) yields \(+0.42 \pm 0.04\%e\) relative to NIWA (not shown in Fig. 2a), but is inconsistent with the above intercomparison results. The determinate error has yet to be resolved.

5.5.2 \(\delta^2\text{D-Ch}_4\)

According to the intercomparison in this study (Sect. 4.1), the UCI has a \(\delta^2\text{D-Ch}_4\) offset of \(-10.7 \pm 0.7\%e\) with respect to IMAU (no. 5 in Fig. 2b).

5.6 TU

5.6.1 \(\delta^13\text{C-Ch}_4\)

The intercomparison in this study (Sect. 3.2) and the IMAU offset (Sect. 5.2) give an offset of the TU measurements relative to NIWA to be \(+0.36 \pm 0.08\%e\) (no. 6 in Fig. 2b). Measurements at TU have been regularly compared with those at NIPR and they are in agreement within reproducibility of both systems (Umezawa et al., 2009 and additional measurements since then). This is consistent with the previous intercomparison between NIPR and NIWA (Sect. 5.7) and indicates long-term intra-laboratory consistency of TU and NIPR measurements. It is reasonable that TU shares the offset level with NIPR, because both institutions use the same standard scale. As described in Sect. 2.6, it should be noted that the above offset value is not for the data sets currently available to the research community (Umezawa et al., 2011, 2012a, b), for which \(+0.32 \pm 0.08\%e\) (not shown in Fig. 2) is recommended. Correction of the data sets from the earlier publications is under evaluation.

Figure 2. (a) \(\delta^13\text{C-Ch}_4\) offsets of the different laboratories with respect to the NIWA DI-IRMS measurement with grey shades for ease of viewing. (b) \(\delta^2\text{D-Ch}_4\) offsets of the different laboratories with respect to the IMAU GC-IRMS measurement. See Table 1 and text for corresponding subsections in Sects. 3 and 5. Numbers shown in each laboratory column are the plausible measurement offsets estimated in this study. Note that this result represents intercomparisons for the modern atmospheric \text{CH}_4.
The intercomparison in this study (Sect. 4.2) gives an offset of $-13.1 \pm 0.6\%e$ for the TU atmospheric $\delta^{13}$C-CH$_4$ measurement (no. 6 in Fig. 2b).

An intercomparison between NIPR and NIWA was conducted in 2004 (Morimoto et al., 2006). After the recent update of the NIPR standard scale, the NIPR offset is evaluated to be $+0.33 \pm 0.04\%e$ higher than NIWA (top in no. 7, Fig. 2a). The intercomparison in this study (Sect. 4.2) combined with the IMAU offset (Sect. 5.2) indicates that the NIPR measurement is $+0.27 \pm 0.08\%e$ with respect to NIWA (bottom in no. 7, Fig. 2a), which is consistent with the above value.

To our knowledge, no intercomparison exercises with UW have been reported.

Levin et al. (2012) estimated the UHEI $\delta^{13}$C-CH$_4$ offset to be $-0.169 \pm 0.031\%e$ relative to NIWA (top in no. 9, Fig. 2a). The intercomparison between UHEI and MPI-BGC in this study (Sect. 3.3), together with the MPI-BGC offset (Sect. 5.4), also infers the UHEI offset to be $-0.05 \pm 0.13\%e$ (bottom in no. 9, Fig. 2a), which is consistent with the above value. Earlier measurements of three air samples at both UHEI and NIWA indicated that the UHEI offset is $-0.04 \pm 0.04\%e$ relative to NIWA (Poß, 2003; Behrens et al., 2008). It is also noted that, in an intercomparison presented by Nisbet (2005), the UHEI measurement was $-0.07 \pm 0.04\%e$ lower than NIWA. As these earlier comparison results have been published before the rigorous corrections of the UHEI measurements, these values are not included in Fig. 2a.

Levin et al. (2012) estimated the INSTAAR measurements have an offset of $+0.132 \pm 0.022\%e$ with respect to NIWA (top in no. 10, Fig. 2a). In an intercomparison exercise reported by Nisbet (2005), the INSTAAR measurement was $+0.14 \pm 0.06\%e$ higher than NIWA (not shown in Fig. 2a), which is consistent with the above value. This study (Sect. 4.4) indicates that the INSTAAR measurement is $+0.15 \pm 0.05\%e$ higher than NIWA for the cylinder with high CH$_4$ mole fraction (CA 03560; second top of no. 10 in Fig. 2a). The intercomparison between INSTAAR and MPI-BGC (Sect. 4.5) indicates that, combined with the MPI-BGC offset (Sect. 5.4), the INSTAAR offset is $+0.21 \pm 0.12\%e$ relative to NIWA (second bottom in no. 10, Fig. 2a). Lastly, the co-located sample intercomparison (Sect. 4.6) indicates the INSTAAR offset to be $+0.08 \pm 0.11\%e$ (bottom in no. 10, Fig. 2a). It is important to note again that only the round robin intercomparison measurements (Sect. 4.4 and second top of no. 10 in Fig. 2a) and the intercomparison with MPI-BGC (Sect. 4.5) were carried out with a PCS column to remove the Kr interference and that the data set currently available to the public from INSTAAR will be evaluated for future correction.

As described in Sect. 2.10, INSTAAR follows the standard scale of UCI. Tyler et al. (2007) reported that measurements of 10 air cylinders filled at Niwot Ridge, Colorado in 2000–2001 were analysed at both laboratories and that the result indicated an offset of INSTAAR to be $+0.04 \pm 0.12\%e$ relative to UCI. The collection of air samples at Niwot Ridge for the UCI–INSTAAR comparison continued until 2003. A revisit to the measurement record showed that the INSTAAR offset relative to UCI had shifted over the years; the average differences are $+0.02 \pm 0.08$ for 2000 ($N = 7$), $+0.12 \pm 0.07\%e$ for 2001 ($N = 2$) and $+0.26 \pm 0.03\%e$ for 2002 ($N = 12$). This fact may suggest excursions of the internal calibration of either laboratory for these years, but the cause has yet to be resolved; this problem will be addressed in a subsequent paper from either group. It is noted that the offsets relative to NIWA for both laboratories inferred from the different intercomparison pathways are consistent with each other within the uncertainties (Fig. 2a).

Bock et al. (2010) reported an intercomparison between UB and INSTAAR. This indicates that the INSTAAR measurement-
ment offset is $-13.2 \pm 1.3\%e$ with respect to IMAU (no. 10 in Fig. 2b).

5.11 RHUL

5.11.1 $\delta^{13}$C-$\text{CH}_4$

Nisbet (2005) reported that the RHUL DI-IRMS measurements agreed well with NIWA with an offset of $0.00 \pm 0.02\%e$ (top in no. 11, Fig. 2a). At the same time, they indicated that the RHUL GC-IRMS measurement has an offset of $+0.11 \pm 0.13\%e$ with respect to NIWA, and later Nisbet et al. (2016) reported that the GC-IRMS system has an offset of about $+0.3\%e$ relative to NIWA (not shown in Fig. 2a). Based on measurements of air in two cylinders exchanged between RHUL and NIWA in 2011 and 2014, RHUL applied an offset correction ($-0.20\%e$) to all data (see Sect. 4.6), by which the RHUL offset has now been evaluated to be $+0.12 \pm 0.03\%e$ (middle of no. 11 in Fig. 2a). The intercomparisons based on the co-located air samples via INSTAAR (Sect. 4.6), combined with the INSTAAR offset (Sect. 5.10), infer that the RHUL offset is $+0.10 \pm 0.03\%e$ relative to NIWA (bottom in no. 11, Fig. 2a).

5.12 PDX

5.12.1 $\delta^{13}$C-$\text{CH}_4$

Rice et al. (2016) presented an offset of $+0.024 \pm 0.088\%e$ of the PDX measurements relative to UW by comparing coinciding measurements of archive air samples at PDX and $\delta^{13}$C-$\text{CH}_4$ records from Quay et al. (1999) from stations Mauna Loa, Hawaii and Tutuila, American Samoa (1995–1996). With the UW offset with respect to NIWA (Sect. 5.8), it is indicated that the PDX measurement is $+0.08 \pm 0.09\%e$ higher than NIWA (no. 12 in Fig. 2a). This offset is consistent with the UCI offset with respect to NIWA within the uncertainties (note that PDX follows the UCI standard scale).

5.12.2 $\delta^{D-\text{CH}_4}$

Since PDX follows the UCI standard scale (Teama, 2013; Rice et al., 2016), the likely offset is the same as that of UCI (no. 12 in Fig. 2b).

5.13 PSU

5.13.1 $\delta^{13}$C-$\text{CH}_4$

According to Schmitt et al. (2013), the PSU measurement has an offset of $+0.03 \pm 0.16\%e$ relative to NIWA after being corrected for the Kr interference. The measurements of the cylinder with a high CH$_4$ mole fraction (CA 03560) at PSU are $+0.03 \pm 0.16$, $+0.27 \pm 0.16$ and $+0.13 \pm 0.05\%e$ (no. 13 top, middle and bottom, respectively in Fig. 2a) higher than NIWA for different Kr corrections at different measurement times, these values being consistent with each other within the uncertainties.

5.13.2 $\delta^{D-\text{CH}_4}$

An intercomparison result using three firn air samples gives the PSU an offset of $-12.1 \pm 1.5\%e$ relative to the IMAU measurement (Sapart et al., 2011; top in no. 13, Fig. 2b). The intercomparison in this study (Sect. 4.4) gives $-13.6 \pm 1.5\%e$ relative to NIWA for the cylinder with a high CH$_4$ mole fraction (CA 03560); bottom in no. 13, Fig. 2b).

5.14 UB

5.14.1 $\delta^{13}$C-$\text{CH}_4$

The UB measurement has an offset of $-0.18 \pm 0.09\%e$ relative to NIWA (Schmitt et al., 2013; no. 14 in Fig. 2a). This was determined by the round robin intercomparison (Sect. 4.4 and Table 5).

5.14.2 $\delta^{D-\text{CH}_4}$

Sapart et al. (2011) gives an intercomparison result between UB and IMAU, indicating the UB offset of $0.0 \pm 1.6\%e$ relative to IMAU (top in no. 14, Fig. 2b). This value is consistent with the intercomparisons between UB and IMAU reported by Bock et al. (2010). Later UB modified the measurement set-up, but the measurements of same air samples before and after all modifications were in good agreement as presented by Bock et al. (2014). The intercomparison in this study (Sect. 3.4) shows that the UB measurement differs insignificantly by $-0.8 \pm 2.5\%e$ with respect to IMAU for the cylinder with high CH$_4$ mole fraction (CA 03560; bottom in no. 14, Fig. 2b).

5.15 AWI

5.15.1 $\delta^{13}$C-$\text{CH}_4$

The AWI offset is reported to be $-0.09 \pm 0.06\%e$ with respect to NIWA (Schmitt et al., 2013; no. 15 in Fig. 2a).

5.16 CIC

5.16.1 $\delta^{13}$C-$\text{CH}_4$

Sperlich et al. (2012) reported measurements of an air cylinder at CIC, IMAU and UB. The CIC measurement insignificantly different by $+0.01 \pm 0.09\%e$ from IMAU, and the CIC offset with respect to NIWA is estimated to be $-0.03 \pm 0.11$ (top in no. 16, Fig. 2a). They have also reported that the CIC measurement is in agreement with UB with difference of $+0.00 \pm 0.14\%e$. It is noted that, although the UB offset relative to NIWA is estimated to be significant (Sect. 5.14), the difference is still within uncertainties of the intercomparison exercises. Two pure CH$_4$ gases prepared by Sperlich
et al. (2012) constitute crucial components of the reference gas series developed at MPI-BGC (Sperlich et al., 2016). This has provided a direct intercomparison between CIC and MPI-BGC. The CIC measurement is $+0.09 \pm 0.14\%$ higher than MPI-BGC. Combined with the MPI-BGC offset (Sect. 5.4), the CIC offset with respect to NIWA is estimated to be $+0.02 \pm 0.18\%$ (bottom in no. 16, Fig. 2a), which is consistent with the aforementioned value.

### 5.16.2 δ-D-CH$_4$

Sperlich et al. (2016) reported δ-D-CH$_4$ measurement results of the two reference gases prepared by Sperlich et al. (2012) at CIC and MPI-BGC. The results indicated that the CIC measurement differs by $+2.1 \pm 1.8\%$ from MPI-BGC. Combined with the MPI-BGC offset (Sect. 4.4), the CIC offset relative to IMAU is estimated to be $−2.1 \pm 2.1\%$ (no. 16 in Fig. 2b).

### 6 Summary and discussion

We carried out interlaboratory comparison exercises for atmospheric δ$^{13}$C-CH$_4$ and δ-D-CH$_4$ covering many laboratories around the world. In addition, we reviewed previously published intercomparison results. The results indicated measurement offsets among laboratories, which range from $−0.2$ to $+0.3\%$ with respect to the NIWA DI-IRMS measurement for δ$^{13}$C-CH$_4$ and up to $−13\%$ with respect to the IMAU measurement for δ-D-CH$_4$. These offset values are larger than the measurement uncertainties from individual laboratories.

The significant δ$^{13}$C-CH$_4$ measurement offsets among laboratories are obvious even though all laboratories ultimately refer to the VPDB scale. We have presented potential causes of the measurement offsets in individual laboratories (Sect. 2), with possible further causes being hidden in all preparation and measurement steps of standard materials. (1) The scale contraction effect for DI-IRMS CO$_2$ analysis, which is instrument dependent, could be responsible for a considerable part of the observed offsets, given the fact that the atmospheric δ$^{13}$C-CH$_4$ value (about $−47\%$) differs considerably from the primary anchor of the VPDB scale (NBS-19). (2) Individual laboratories have carried out calibrations against different RMs with different uncertainties of assigned values; such diverse calibration trajectories have also definitely contributed to the interlaboratory measurement offsets. Such RMs have different chemical properties and are processed to CO$_2$ at individual laboratories, at which different fractionation is possible. (3) Different algorithms for $^{17}$O correction have been used for δ$^{13}$C measurements at different laboratories, which could have caused biases among available data sets. (4) The Kr interference on a GC-IRMS system is in several cases a probable cause of the offsets, and unfortunately, this effect is system dependent and can vary with time and the instrument settings. Lastly, it is important to note that we summarized δ$^{13}$C-CH$_4$ measurement offsets at the modern atmospheric CH$_4$ mole fraction level, but the offset may vary with the amount of CH$_4$ analysed (e.g. lower mole fractions in ice core analyses, see Tables 3, 5 and 6), because of a non-linear response of IRMS (Umezawa et al., 2009) and because the Kr interference is directly dependent on the Kr-to-CH$_4$ ratio (Schmitt et al., 2013). Furthermore, the intercomparisons presented here focus on modern atmospheric CH$_4$ of typically $−47\%$ and such comparisons for high and low δ$^{13}$C-CH$_4$ values (e.g. CH$_4$ from ice cores or enriched/depleted source signatures) are to date very limited (Tables 3 and 6 in this study).

Concerning δ-D-CH$_4$ measurement offsets among laboratories, it is interesting that the listed laboratories can be roughly split into two groups whose δ-D-CH$_4$ measurements differ by $\sim 10\%$. Some laboratories with higher δ-D-CH$_4$ values refer to an identical set of standards produced at the MPIC (MPIC and IMAU) or to the UHEI calibration (UHEI and UB), and measurements of these groups have been cross-referenced (see Sects. 2 and 4), thereby showing the reasonable agreements. The original calibrations were carried out using an offline CH$_4$ processing technique (cryogenic separation and conversion of CH$_4$ to CO$_2$ and H$_2$O followed by H$_2$O reduction to H$_2$) with subsequent analysis by DI-IRMS. The other laboratories with higher δ-D-CH$_4$ values recently developed their own primary calibrations independently (CIC and MPI-BGC). CIC used an offline CH$_4$ processing technique combined with DI-IRMS, whereas MPI-BGC adopted TC/EA coupled to continuous-flow IRMS. For the lower δ-D-CH$_4$ group, some laboratories carried out calibrations against Oztech H$_2$ gases (UCI, PDX and PSU) or have other calibration pathways (TU and INSTAAR; see Sect. 2). These laboratories used local H$_2$ working gas standards for GC-IRMS, which were calibrated with a separate DI-IRMS procedure. As is the case for δ$^{13}$C-CH$_4$, possible causes of the observed δ-D-CH$_4$ discrepancies could have arisen in all preparation and measurement steps. (1) The classical technique for DI-IRMS involves processing of H$_2$O, and the associated steps in experimental lines are prone to surface adhesion and contamination of H$_2$O; thereby considerable memory effect is possible (Bergmaschi et al., 2000). (2) Similarly to δ$^{13}$C-CH$_4$, calibration for δ-D-CH$_4$ involves measurements of standards with different chemical properties (H$_2$O and H$_2$), and such calibrations at different laboratories could contribute to the offset. (3) Difficulties in maintaining stable pyrolysis conditions for GC-IRMS (Bock et al. 2010) might have affected measurements against local H$_2$ working standards in the cases where the principle of identical treatment (Werner and Brand, 2001) was not strictly followed. Lastly, it is noted that the non-linearity of the IRMS in δ-D-CH$_4$ measurements (Brass and Röckmann, 2010) may also play a role for samples with low mole fractions such as ice core analyses.
The measurement offsets summarized in this study should be thoroughly taken into account when data from different laboratories are combined, and this study will be of help when incorporating merged $\delta^{13}$C-CH$_4$ and $\delta^D$-CH$_4$ data sets into a state-of-the-art chemistry transport model. However, it is recommended that data users contact the data providers directly for the latest information whenever possible. The Kr interference is under evaluation at some laboratories and it will possibly involve an update of the data sets that are currently available. More importantly, it is imperative to have common reference gases with transparent and reproducible traceability (for instance, Sperlich et al. 2016) and to carry out a systematic intercomparison programme (flask or cylinder round robin) in the research community to attain the necessary but ambitious high-compatibility goals of 0.02 ‰ for $\delta^{13}$C-CH$_4$ and 1 ‰ for $\delta^D$-CH$_4$ (WMO, 2016). Such thorough efforts will facilitate optimized use of $\delta^{13}$C-CH$_4$ and $\delta^D$-CH$_4$ data sets in a combined way and maximize the number of isotope data sets (and thus their spatial and temporal coverage) usable for enhancing our understanding of the global CH$_4$ cycle.

We welcome collaborative works that analyse the multiple data sets from laboratories that participated in this study (see data availability listed in Table 1). Data users can examine the offset numbers (Table 1 and Fig. 2) and adjust the data sets at least for data points with values close to the modern atmosphere in $\delta^{13}$C-CH$_4$ and $\delta^D$-CH$_4$ as well as the CH$_4$ mole fraction. For data with CH$_4$ mole fractions and isotopic ratios that are far from modern background values (e.g. sample air from ice core and stratosphere and those influenced by sources), more intercomparisons are needed to establish correction factors among data sets.

Data availability. All the interlaboratory comparison data presented in this study are included in the tables of this paper.
Appendix A: List of participating institution/project acronyms

AWI: Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
CARIBIC: Civil Aircraft for the Regular Investigation of the atmosphere Based on an Instrument Container
CIC: Centre for Ice and Climate, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
IMAU: Institute for Marine and Atmospheric research Utrecht, Utrecht University, Utrecht, the Netherlands
INSTAAR: Institute of Arctic and Alpine Research University of Colorado Boulder, Boulder, USA
MPI-BGC: Max Planck Institute for Biogeochemistry, Jena, Germany
MPIC: Max Planck Institute for Chemistry, Mainz, Germany
NCAR: National Center for Atmospheric Research, Boulder, USA
NIPR: National Institute of Polar Research, Tokyo, Japan
NIWA: National Institute for Water and Atmospheric Research, Wellington, New Zealand
NOAA: National Oceanic and Atmospheric Administration, USA
PDX: Portland State University, Portland, USA
PSU: Pennsylvania State University, Pennsylvania, USA
RHUL: Royal Holloway, University of London, Egham, UK
TU: Tohoku University, Sendai, Japan
UB: University of Bern, Bern, Switzerland
UCI: University of California Irvine, Irvine, USA
UHEI: University of Heidelberg, Heidelberg, Germany
UW: University of Washington, Seattle, USA
Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We thank Morimasa Sato for measurements at NIES. We thank Paul Quay (UW) for isotopic CH$_4$ from co-located samples measured by INSTAAR, RHUL and NIWA. We thank Paul Quay (UW) for isotopic CH$_4$ data and intercalibration information and Doaa Teama for isotopic measurements at PDX. We thank Célia Sapart for helpful information about the past intercomparison. Part of this work at NIES was supported by the Environment Research and Technology Development Fund (2-1710) of the Ministry of the Environment, Japan and Environmental Restoration and Conservation Agency. Funding for work at PDX was provided by the US National Science Foundation (Atmospheric and Geospace Sciences Grant 0952307). CH$_4$ isotope ice core work at UB received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) ERC grant agreement no. 226172 and the Swiss National Science Foundation (grant no. 200020_159563 & 20020_172506). NIWA's isotope measurements are funded under the Climate and Atmosphere Research Programme CAAC1804 (2017/18 SCI). We are grateful to the referee Sergey Assonov and one anonymous referee for their insightful and constructive comments for improving this paper.

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

Edited by: Frank Keppler
Reviewed by: Sergey Assonov and one anonymous referee

References


T. Umezawa et al.: Interlaboratory comparison of $\delta^{13}$C and $\delta$D measurements of CH$_4$


