

Harmonisation of methane isotope ratio measurements from different laboratories using atmospheric samples

Bibhasvata Dasgupta¹, Malika Menoud¹, Carina van der Veen¹, Ingeborg Levin^{2,†}, Cordelia Veidt², Heiko Moossen³, Sylvia Englund Michel⁴, Peter Sperlich⁵, Shinji Morimoto⁶, Ryo Fujita^{6,7}, Taku Umezawa^{6,8}, Stephen Platt⁹, Christine Groot Zwaafink⁹, Cathrine Lund Myhre⁹, Rebecca Fisher¹⁰, David Lowry¹⁰, Euan Nisbet¹⁰, James France¹⁰, Ceres Woolley Maisch^{1,10}, Gordon Brailsford⁵, Rowena Moss⁵, Daisuke Goto¹¹, Sudhanshu Pandey¹², Sander Houweling¹³, Nicola Warwick¹⁴, Thomas Röckmann¹

¹Institute for Marine and Atmospheric research Utrecht (UU), Utrecht University, The Netherlands

²Institut für Umweltphysik, Heidelberg University, INF 229, 69120 Heidelberg, Germany

³Max-Planck-Institute for Biogeochemistry (BGC-IsoLab), Jena, Germany

⁴Institute of Arctic and Alpine Research (INSTAAR), University of Colorado, Boulder, USA

⁵National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand

⁶Center for Atmospheric and Oceanic Studies, Graduate School of Science, Tohoku University, Sendai, Japan

⁷Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan

⁸National Institute for Environmental Studies, Tsukuba, Japan

⁹NILU, P.O. Box 100, 2027 Kjeller, Norway

¹⁰Centre of Climate, Ocean and Atmosphere, Department of Earth Sciences, Royal Holloway, University of London, Egham, UK

¹¹National Institute of Polar Research, Tokyo, Japan

¹²Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

¹³Department of Earth Sciences, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

¹⁴Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, UK

[†]deceased

Correspondence to: Bibhasvata Dasgupta (bdasgupta03@gmail.com; bdasgupta@uu.nl)

Abstract. Establishing interlaboratory compatibility among measurements of stable isotope ratios of atmospheric methane ($\delta^{13}\text{C}-\text{CH}_4$ and $\delta\text{D}-\text{CH}_4$) is challenging. Significant offsets are common because laboratories have different ties to the VPDB or SMOW-SLAP scales. Umezawa et al. (2018) surveyed numerous comparison efforts for CH_4 isotope measurements conducted from 2003 to 2017 and found scale offsets of up to 0.5 ‰ for $\delta^{13}\text{C}-\text{CH}_4$ and 13 ‰ for $\delta\text{D}-\text{CH}_4$ between laboratories.

This exceeds the World Meteorological Organisation Global Atmospheric Watch (WMO-GAW) network compatibility targets of 0.02 ‰ and 1 ‰ considerably.

We employ a method to establish scale offsets between laboratories using their reported CH_4 isotope measurements on atmospheric samples. Our study includes data from eight laboratories with experience in high-precision isotope ratio mass spectrometry (IRMS) measurements for atmospheric CH_4 . The analysis relies exclusively on routine atmospheric measurements conducted by these laboratories at high-latitude stations in the Northern and Southern Hemispheres, where we assume each measurement represents sufficiently well-mixed air at the latitude for direct comparison. We use two methodologies for interlaboratory comparisons: (I) assessing differences between time-adjacent observation data and (II) smoothing the observed data using polynomial and harmonic functions before comparison. The results of both methods are consistent, and with a few exceptions, the overall average offsets between laboratories align well with those reported by Umezawa et al. (2018). This indicates that interlaboratory offsets remain robust over multi-year periods. The evaluation of routine measurements allows us to calculate the interlaboratory offsets from hundreds, in some cases thousands of measurements. Therefore, the uncertainty in the mean interlaboratory offset is not limited by the analytical error of a single analysis but by real atmospheric variability between the sampling dates and stations. Using the same method, we assess this uncertainty by investigating measurements from four high-latitude sites analysed by the INSTAAR laboratory. After applying the derived interlaboratory offsets, we present a harmonised time series for $\delta^{13}\text{C}-\text{CH}_4$ and $\delta\text{D}-\text{CH}_4$ at high northern and southern latitudes, covering the period from 1988 to 2023.

Methane (CH₄) is a potent greenhouse gas, with a global warming potential 28 times larger than carbon dioxide (CO₂) over a 100-year time scale (Dean et al., 2018). It is released from both natural sources, such as wetlands, geological seepage, termites, and natural forest fires, and anthropogenic sources, such as agriculture, waste, and fossil fuel extraction and use (Kirschke et al., 2019; Saunio et al., 2016). The isotopic composition of CH₄, specifically the ratios of ¹³C to ¹²C (expressed as a relative deviation from a standard, δ¹³C-CH₄) and ²H to ¹H (expressed as δD-CH₄), provides valuable information about its sources and sinks (Quay et al., 1999; Bréas et al., 2001; Brenninkmeijer et al., 2003; Whiticar, 2020; Michel et al., 2024). Harmonising CH₄ isotope data between measurements carried out at different laboratories is crucial for accurately interpreting recent trends in global CH₄ growth. This is particularly relevant to current climate policy priorities, such as the Global Methane Pledge, which aims to reduce anthropogenic CH₄ emissions by 30 % in 2030 compared to 2020. Unambiguously disentangling anthropogenic from natural sources will be required to effectively evaluate progress toward meeting the pledge's goals.

Each source type emits CH₄ with a characteristic isotopic signature range, which is influenced by the isotopic composition of the substrates it is formed from, as well as the conditions and processes involved in CH₄ formation. Similarly, the processes that remove CH₄ from the atmosphere, primarily oxidation by the hydroxyl radical (OH), with smaller contributions from chlorine (Cl), electronically excited oxygen atoms (O(¹D)) and soil microbes, are associated with specific kinetic isotope effects, leading to changes in the isotopic composition of the remaining atmospheric CH₄ (Fujita et al., 2020; Brenninkmeijer et al., 2003; Whiticar, 2020; Röckmann et al., 2011). The global isotopic composition of atmospheric CH₄ reflects the combined isotopic signatures of its sources, adjusted for the kinetic isotope effects of removal processes (Quay et al., 1999; Whiticar, 2020). Changes in the isotopic composition can indicate shifts in the source mix, variations in removal rates, or imbalances between sources and sinks.

High-precision isotopic measurements have shown a reversal in the δ¹³C-CH₄ trend over the past decades. These isotope signals have been used to constrain the global CH₄ budget and to investigate changes in the relative contributions of different sources over time (e.g., Bousquet et al., 2006; Monteil et al., 2011; Schaefer et al., 2016; Nisbet et al., 2016; 2023; Schwietzke et al., 2016; Worden et al., 2017; Basu et al., 2022; Lan et al., 2021; Thanwerdas et al., 2024; Chandra et al., 2024; Michel et al., 2024). The generally emerging picture is that the isotope trend reflects a shift in the source mix from isotopically enriched fossil and combustion sources to isotopically depleted biogenic sources, although changes in the sink and temporal trends in the isotopic composition of each source type can also contribute to the shift (Rigby et al., 2017; Turner et al., 2017; Thanwerdas et al., 2024). These conclusions would have been difficult to reach without isotopic analyses.

Despite instrumental advancements in CH₄ isotope measurements and careful evaluation of experimental procedures, significant measurement offsets between laboratories exist. Such offsets are primarily due to different realisations of the reference scales in laboratories, but can also arise from artefacts in the analytical methods. High-precision CH₄ isotope measurements are mostly performed using Isotope-Ratio Mass Spectrometry (IRMS), although laser-based methods are gaining importance (Bergamaschi et al., 1994; Keppler et al., 2010; Eyer et al., 2016; Röckmann et al., 2016; Rennick et al., 2021), adding another layer of variability. We focus on IRMS measurements here. Each laboratory has protocols for purifying CH₄ from other atmospheric gases and converting it to CO₂ or hydrogen (H₂) to determine the δ¹³C-CH₄ or δD-CH₄ values, respectively. These values are then reported relative to the international primary reference scales: VPDB (Vienna PeeDeeBelemnite; defined by carbonate as reference material) for δ¹³C-CH₄ and VSMOW/SLAP (Vienna Standard Mean Ocean Water/Standard Light Antarctic Precipitation; defined by water as reference material) for δD-CH₄. However, the reference materials (RMs) that link these scales are isotopically and materially different, meaning their use application has

compositions and physical properties that do not match atmospheric CH₄.

Some laboratories have invested significant efforts to link their internal CH₄-in-air standards to the primary RMs. This is done through careful experiments that convert carbon and hydrogen in CH₄ into other analytes (Sperlich et al., 2012; 90 2016; Morimoto et al., 2017; Sperlich et al., 2020). Other laboratories align their reference scales by propagating their working standards from established laboratories, creating ‘families’ of laboratories with scale differences relative to other ‘families’ (Umezawa et al., 2018). Analytical artefacts can occur in CH₄ measurements when the *principle of identical treatment* between sample and reference material is not followed (Werner & Brand, 2001), but also due to differences in the ¹⁷O correction (Assonov and Brenninkmeijer, 2003) or mass interferences from contaminants such as Krypton (Schmitt et al., 2013). 95 Establishing scales for high-precision CH₄ measurements is therefore challenging and requires expert laboratories to employ robust referencing and quality control (QC) and quality assurance (QA) strategies to ensure accuracy (see Umezawa et al., 2018 for an overview).

Occasionally, laboratories engage in bilateral or multi-laboratory ‘round-robin’ (RR) exercises where several gas mixtures are circulated and measured, and the results are compared. They can be specifically designed to provide challenging 100 matrices, addressing isobaric contamination issues like Krypton in CH₄ measurements (Schmitt et al., 2013). RR exercises are invaluable because they allow direct comparison of laboratory measurements on identical samples, ensuring consistency and reliability and focusing solely on the laboratory’s instrumental and data processing capabilities. Another comparison method uses samples from co-located stations sampling a homogenised atmosphere, which may introduce variability due to differences in sample collection times, locations, procedures and containers, complicating interlaboratory consistency (Steuer et al., 2023). 105 Umezawa et al. (2018) conducted a comprehensive survey on the isotopic composition of CH₄, consolidating prior interlaboratory comparisons to evaluate offsets among 17 laboratories. Their study, which included offsets from RRs and co-located samples, provided a large-scale assessment of data consistency across laboratories measuring isotopes in atmospheric CH₄. However, interlaboratory offsets may have shifted, especially as δ D-CH₄ measurements have become more routine. Additionally, Umezawa et al. (2018) reported offsets were derived from various smaller intercomparison subsets rather than a 110 single, unified RR, warranting further investigation.

In this study, we compile, evaluate, offset-correct, and merge the time series of CH₄ mole fraction and isotopic composition measured in eight different laboratories on ambient air samples collected at high-latitude stations from the Northern Hemisphere (NH) and Southern Hemisphere (SH) (Table 1). The results of our air-sample-based data harmonisation approach are compared to the offsets reported by Umezawa et al. (2018) and RR exercises. Once the offsets are established, 115 all the time series are transferred to a common scale before being merged within each hemisphere to create continuous, high-latitude, hemisphere-specific time series for each isotope signature.

2. Methods

2.1. Datasets

The CH₄ mole fraction, $\delta^{13}\text{C-CH}_4$ and $\delta\text{D-CH}_4$ datasets were obtained from 10 air sampling stations: six in the NH 120 and four in the SH, provided by eight research groups that carry out long-term measurements of CH₄ isotopic composition: the Max Planck Institute for Biogeochemistry (MPI-BGC) in Jena, Germany, the Institute of Arctic and Alpine Research (INSTAAR) in Boulder, Colorado USA (collected in cooperation with the NOAA Global Monitoring Laboratory), the National Institute of Water and Atmospheric Research Ltd. (NIWA) in Wellington, New Zealand, Tohoku University and National

of London (RHUL) in the UK and Utrecht University (UU) in the Netherlands. The datasets collected are listed in Table 1, and Fig. 1 shows the temporal coverage of the two isotopic tracers in both hemispheres from each participating laboratory.

The MPI-BGC laboratory has conducted extensive efforts in the past years to link their scale to the primary international RMs (Brand et al., 2016; Sperlich et al., 2016; Sperlich et al., 2020). Therefore, the MPI-BGC dataset is a suitable reference laboratory for the comparison. However, MPI-BGC's high-latitude measurements only started in 2011 for NH sites and in 2018 for SH sites; thus, the overlap period with other, longer time series is limited. Umezawa et al. (2018) used NIWA as a reference laboratory for $\delta^{13}\text{C}-\text{CH}_4$, but NIWA has not analysed air from high-latitude sites in the NH, whereas most other laboratories have done so. To extend the comparison period between laboratories beyond the timescale covered by MPI-BGC, we have also used TU/NIPR as an alternative reference laboratory. TU/NIPR has reported $\delta^{13}\text{C}-\text{CH}_4$ and $\delta\text{D}-\text{CH}_4$ data from high-latitude stations in NH since 1996 and SH since 1995, respectively. The dual approach also allows us to examine whether the harmonisation of datasets depends on the choice of the reference laboratory.

Hemisphere	Station	Latitude (°)	Longitude (°)	Altitude (m)	Laboratory	$\delta^{13}\text{C}-\text{CH}_4$				Reference	$\delta\text{D}-\text{CH}_4$				Reference
						Observation Period	Number of Observations	Sampling Frequency (per month)	Measurement Uncertainty (‰)		Observation Period	Number of Observations	Sampling Frequency (per month)	Measurement Uncertainty (‰)	
NH	Zeppelin (ZEP)	78.9067	11.8883	475	INSTAAR	2008.45 to 2023.99	1122	6.02	0.10	Lan et al., 2021	2008.31 to 2009.82	19	1.05	2.0	White et al., 2016
					RHUL	2012.52 to 2024.02	1782	12.91	0.05	Fisher et al., 2005; Woolley Maisch et al., 2023					
					UU	2023.16 to 2023.64	29	5.04	0.07	Röckmann et al., 2016	2012.71 to 2023.64	573	4.37	2.0	Röckmann et al., 2016
	Ny-Alesund (NYA)	78.9235	11.9236	475	TU/NIPR	1996.18 to 2022.97	1187	3.69	0.08	Tanaka et al., 1987; Morimoto et al., 2006; Fujita et al., 2018, 2020	2005.66 to 2023.67	800	3.70	2.2	Tanaka et al., 1987; Morimoto et al., 2006; Fujita et al., 2018, 2020
	Summit (SUM)	72.5962	-38.422	3238	INSTAAR	2010.32 to 2023.59	918	5.77	0.10	Lan et al., 2021					
	Alert (ALT)	82.4508	-62.5072	200	MPI-BGC	2011.05 to 2023.47	240	1.61	0.12	Brand et al., 2016	2011.05 to 2023.47	230	1.54	1.0	Brand et al., 2016
					UHEI	1993.90 to 2003.93	120	1.00	0.09	Levin et al., 2012	1992.09 to 2003.93	111	0.78		
					INSTAAR	2000.61 to 2023.89	2031	7.27	0.10	Lan et al., 2021	2005.30 to 2009.94	200	3.59	2.0	White et al., 2016
					RHUL	2008.10 to 2019.43	373	2.74	0.05	Fisher et al., 2005; Woolley Maisch et al., 2023					
					UU	2018.91 to 2019.43	25	4.00	0.07	Röckmann et al., 2016	2018.91 to 2019.43	27	4.32	2.0	Röckmann et al., 2016
	Barrow (BRW)	71.32	-156.6	11	INSTAAR	1998.01 to 2024.00	2860	9.17	0.10	Lan et al., 2021	2005.25 to 2010.19	92	1.55	2.0	White et al., 2016
	Kjellnes (KJN)	70.85	29.23	0	MPI-BGC	2014.76 to 2021.31	164	2.08	0.12	Brand et al., 2016	2014.76 to 2021.31	163	2.07	1.0	Brand et al., 2016
SH	Neumayer (NEU)	-70.6666	8.2667	42	MPI-BGC	2018.00 to 2023.99	186	3.11	0.12	Brand et al., 2016	2018.00 to 2022.99	187	3.12	1.0	Brand et al., 2016
					UHEI	1988.49 to 2019.90	300	0.80	0.09	Levin et al., 2012	1988.49 to 2019.79	252	0.67		
					UU	1988.56 to 2020.00	161	0.43	0.07	Röckmann et al., 2016	1988.56 to 2020.00	163	0.43	2.0	Röckmann et al., 2016
	Syowa (SYO)	-69	-39.575	16	TU/NIPR	1995.17 to 2023.06	1140	3.41	0.08	Tanaka et al., 1987; Morimoto et al., 2006; Fujita et al., 2018, 2020	1995.17 to 2022.99	162	0.49	2.2	Tanaka et al., 1987; Morimoto et al., 2006; Fujita et al., 2018, 2020
	Arrival Heights (ARH)	-77.8335	166.6567	189	NIWA	1988.93 to 2022.80	361	0.89	0.02	Schaefer et al., 2016					
	South Pole (SPO)	-89.98	-24.8	2810	INSTAAR	1998.00 to 2023.83	2433	7.85	0.10	Lan et al., 2021	2005.11 to 2009.73	107	1.93	2.0	White et al., 2016

Table 1: List of datasets collected for this study. For TU/NIPR, the $\delta^{13}\text{C}-\text{CH}_4$ data are measured at NIPR and the $\delta\text{D}-\text{CH}_4$ data are measured at TU, respectively.

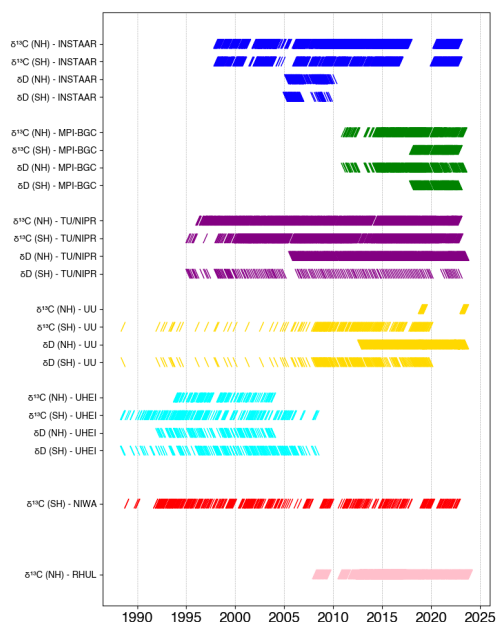


Figure 1: Overview of periods where $\delta^{13}\text{C}\text{-CH}_4$ and $\delta\text{D}\text{-CH}_4$ measurements are available from the eight participating laboratories at high-latitude stations per hemisphere. Note that the measurements from UU for SH samples were carried out on archived air cylinders from UHEI; thus, both laboratories analysed the same air. UHEI analysis was carried out close to sampling dates, but the UU analysis on the archived cylinders in 2021-2022, so years to decades later. Similarly, the $\delta\text{D}\text{-CH}_4$ measurements by TU/NIPR from the SH air were also performed on archived air sample cylinders.

2.2. Data processing

To determine intra-hemispheric interlaboratory offsets, we first combined long-term isotope measurements from all high-latitude sites into laboratory-specific datasets for each hemisphere (NH or SH), and then compared the results between laboratory pairs within each hemisphere. Our approach used quality-controlled data, including any measurement corrections, sourced from the World GHG Data Centre (<https://gaw.kishou.go.jp/>) or provided through personal communication. Each dataset consists of a site description, sampling date, CH_4 mole fraction, isotope composition values, and quality flags for every flask sampled and measured by the laboratory. We compared measurements from stations within the same hemisphere where the reference and test laboratories had data. For example, in the NH, measurements from the MPI-BGC (reference laboratory) at Alert and Kjoelnes stations were compared with those from INSTAAR (test laboratory) at Alert, Barrow, Summit, and Zeppelin stations. In the SH, measurements from MPI-BGC at Neumayer are compared with those from INSTAAR at the South Pole. Figure 2 provides an overview of the data processing used in this study to determine offsets between laboratories. Comparisons are carried out using the nearest neighbour approach (section 2.2.1) or on smoothed fits to the observed data we generated in the study (section 2.2.2). The overlapping period of air sampled between the test and the reference laboratory is considered in both cases. The overlap period between the test and reference laboratory datasets and the sampling frequencies realised over that period determine the total number of data points available for comparison. A longer overlap period and a higher sampling frequency increase the number of comparison points, presumably leading to a more stable and accurate offset calculation, while a shorter overlap period limits the dataset and may introduce greater uncertainty in the final offset. In the

Deleted: compared long-term isotope measurements from high-latitude stations performed by each laboratory in both hemispheres

calculated via an intermediate laboratory to which overlap exists between both laboratories, which in this case is TU/NIPR.

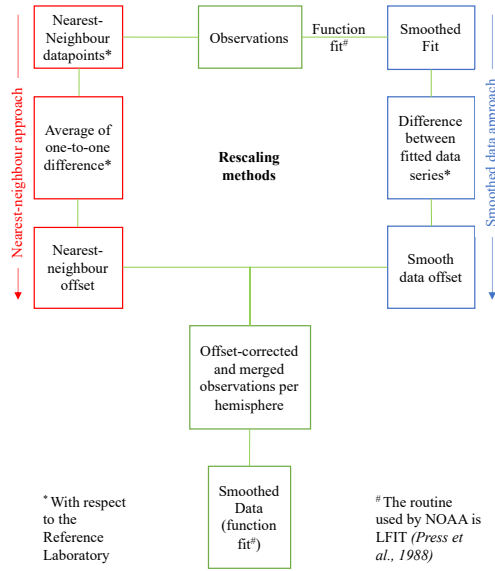


Figure 2: Data processing steps for evaluating, offsetting, and merging the $\delta^{13}\text{C}\text{-CH}_4$ and $\delta\text{D}\text{-CH}_4$ time series compiled from eight laboratories measuring air samples from 10 stations in high northern and southern latitudes.

2.2.1. Nearest-Neighbour approach

In the nearest-neighbour approach, we synchronise the test laboratory's dataset with the reference laboratory's dataset using a nearest-neighbour filter. This process aligns the timeframes by identifying the closest corresponding data point from the reference laboratory within a specified time window for each data point from the test laboratory. The time window is a critical component of this approach, as demonstrated by Levin et al. (2012), who showed good agreement in $\delta^{13}\text{C}\text{-CH}_4$ for air samples taken at the same time across Antarctica. This study applies the same concept. To assess the sensitivity of our results to the choice of the time window, we conducted a series of tests using different time windows (1, 3, 7, 14, and 30 days). The results showed that shorter windows (1-3 days) resulted in fewer matched data pairs but provided more precise alignment with reduced atmospheric variability. Conversely, longer windows (14-30 days) increased the number of matched pairs but introduced greater variability, because these samples are more likely to be representative of different air masses. The final evaluation uses a 7-day window. This is long enough to retain sufficient data pairs between two stations, yet short enough to minimise biases from temporal atmospheric trends.

The process of finding nearest-neighbours means that some data points of the test or reference laboratory may not be used (when a data point from the test laboratory is not a nearest-neighbour to any data point from the reference laboratory) or may be used multiple times (when one datapoint from the test laboratory is nearest-neighbour for several data points from the reference laboratory). This depends on the temporal sampling of the two laboratories being compared. An advantage of the nearest-neighbour approach is that the temporal resolution used to calculate the offset is similar for both laboratories. The disadvantage lies in some data points not being used and others being used multiple times, which could potentially skew the results if those points are outliers or not representative of the background air. The *average difference* from all the nearest-neighbour data pairs returns the mean offset between the two laboratories in the nearest-neighbour approach. The uncertainty

estimated spatial uncertainty using the square root of the sum of their squares.

$$Uncertainty_{nearest-neighbour} = \sqrt{Bootstrap_{error}^2 + Spatial_{uncertainty}^2}$$

A bootstrap is a statistical method that involves repeatedly sampling from a dataset with replacement to estimate the distribution of a statistic. Here, the bootstrap error associated with the mean offset value is derived by randomly resampling the offsets between each reference-test laboratory pair, with the number of resampled datasets set to 10 times the number of data pairs, and computing the 99% confidence interval (CI) from the distribution of resampled means. A spatial uncertainty of 0.06 ‰ for $\delta^{13}\text{C-CH}_4$ and 0.5 ‰ for $\delta\text{D-CH}_4$ is incorporated to account for variations in high-latitude air masses. This estimate is derived from the observed variability among four INSTAAR stations in the NH (section 3.3). This combined uncertainty metric captures the statistical sampling and measurement uncertainty as well as real-world atmospheric differences that could influence tracer measurements.

2.2.2. Smoothed data approach

The smoothed data approach involves fitting the NOAA GML Curve Fitting Method (Thoning et al., 1989) to the observed data of the reference and test laboratories. This fit function captures the annual oscillation, depicted by harmonic cycles, and the long-term growth, represented by a polynomial function, using general linear least squares regression methods (Press et al., 1988).

$$f(t) = a_0 + a_1t + a_2t^2 + a_3t^3 + \sum_{n=1}^4 c_n \sin(2n\pi t + \phi_n) \quad (1)$$

where a_0 , a_1 , a_2 , and a_3 describe the polynomial part of the function, the coefficients c_n describe the harmonic coefficients that determine the amplitude of the seasonal cycles, and the coefficients ϕ_n describe the phase shift for each harmonic, accounting for the timing of seasonal peaks and troughs.

The NOAA CCGCRV method (<https://gml.noaa.gov/aftp/user/thoning/ccgcrv/>) is widely used for decomposing time series data into trend, seasonality, and residuals (Miller and Tans, 2003; Ballantyne et al., 2010; Woolley Maisch et al., 2023). This routine uses a Fast Fourier Transform (FFT) to transform the residuals of the time series fit (with a polynomial trend fit and a harmonic seasonal fit) into the frequency domain. A short-term cutoff filter is first applied to remove the shortest-period components. Long-term and median-term residuals are separated by applying a long-term cutoff to the remaining residuals. The smooth fit for the trend and seasonality is achieved by adding these residuals to the polynomial and harmonic functions, respectively. The final smooth fit of the observed time series is obtained by combining these components. We used the default settings for all parameter thresholds of the NOAA curve fitting method. Specifically, we employ a third-order polynomial function to represent the long-term trend and a fourth-order harmonic function to represent the mean seasonality. For the short-term and long-term cutoff values, we used 80 and 667 days, respectively, which have been shown to be reasonable in previous studies (Dlugokencky et al., 1994; Parker et al., 2018). This means residuals with periods shorter than 80 days are removed from the smooth fit, and a period of 667 days separates seasonal residuals from long-term residuals.

In the smoothed data approach, the offset between the two laboratories is the *average difference* between the two smoothed data series. In practice, this can be easily determined by calculating the long-term average from each laboratory's fitted data and then taking the difference between these two long-term averages. This offset value provides a second estimate of the systematic difference between the two datasets while minimising the impact of short-term fluctuations. The advantage of the smoothed data method is that, when comparing stations with asynchronous sampling times, all the data points contribute equally to the curve fit and, thus, to the offset determination. However, smoothing can sometimes lead to interpolation beyond actual measurement periods, which may introduce discrepancies. For instance, in UU $\delta^{13}\text{C-CH}_4$ measurements for the NH, the

gaps longer than six months, since the NOAA fit becomes unreliable over very sparse data. In this case, the interpolated parts of the curve fit covering significant gaps in data have been excluded from the comparison. When smoothing can be performed without substantial interpolation, both observed and smoothed datasets produce similar offsets and can be used interchangeably within their respective uncertainties. The uncertainty of the smoothed offset is estimated using the square root of the average sum of squares of the root mean square error (RMSE) for the reference and test laboratories' fitted time series, which quantifies the deviation of the observed data from NOAA's fitted curves, i.e., its residual.

$$Uncertainty_{smoothed} = \sqrt{RMSE_{reference}^2 + RMSE_{test}^2}$$

The RMSE is computed as the square root of the average of the squared residuals, providing a measure of the overall fit accuracy and accounting for the spatial uncertainty between stations. This method is chosen as the uncertainty metric because it captures both the variability in the data and the performance of the smoothing model, ensuring that the uncertainty reflects systematic deviations rather than random noise alone.

2.2.3. Comparison to Umezawa et al. (2018)

Umezawa et al. (2018) reported the interlaboratory offsets amongst 16 laboratories for $\delta^{13}\text{C-CH}_4$ and 12 laboratories for $\delta\text{D-CH}_4$, relative to NIWA for $\delta^{13}\text{C-CH}_4$ and UU for $\delta\text{D-CH}_4$ scales, respectively. Eight of those laboratories, including their reference laboratories NIWA and UU, are included in our analysis. For both isotopes, we first converted the offsets from Umezawa et al. (2018) to the scale of our reference laboratories: MPI-BGC and TU/NIPR. To transfer to the MPI-BGC scale, we subtracted the MPI-BGC vs. NIWA (for $\delta^{13}\text{C-CH}_4$) and MPI-BGC vs. UU (for $\delta\text{D-CH}_4$) offsets from the offsets reported for the other laboratories. Similarly, to transfer to the TU/NIPR scale, we subtracted the TU/NIPR-NIWA (for $\delta^{13}\text{C-CH}_4$) and TU/NIPR-UU (for $\delta\text{D-CH}_4$) offsets from the offsets reported for the other laboratories. The converted offsets are shown as grey bars in Fig. 3 with MPI-BGC as the reference laboratory and in Fig. S1 with TU/NIPR as the reference laboratory. In two cases, we adopted different offset values from those recommended by Umezawa et al. (2018) to improve consistency between the two reports. For UHEI vs. NIWA, we selected the -0.04 ± 0.04 ‰ offset reported in Behrens et al. (2008), derived from an RR exercise, rather than the -0.169 ± 0.031 ‰ from Levin et al. (2012), based on co-located sampling at Neumayer. Similarly, for UU vs. NIWA, we used the -0.08 ± 0.11 ‰ offset from Umezawa et al. (2018) over the -0.04 ± 0.03 ‰ from Schmitt et al. (2013). These selections are further motivated and discussed in Section 4.1. The effect of these choices is to change the grey bars to the black-outlined white bars in Fig. 3 and Fig. S1.

Deleted: extrapolated

Deleted: the

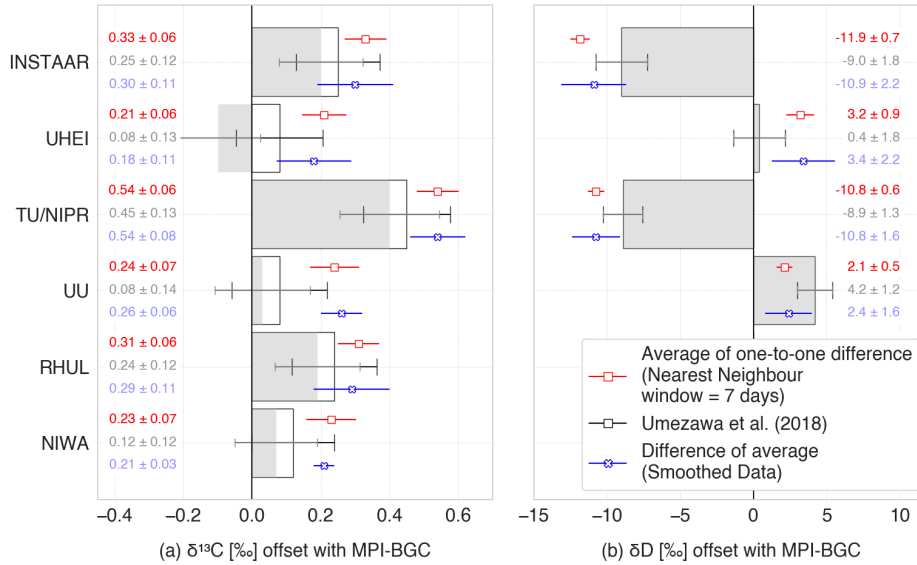


Figure 3: Interlaboratory offsets for (a) $\delta^{13}\text{C}-\text{CH}_4$ and (b) $\delta\text{D}-\text{CH}_4$ with respect to MPI-BGC. Grey bars indicate rescaled Umezawa et al. (2018) values without the modifications mentioned in Section 2.2.3, and black outlined bars show the results after these modifications. The original results from Umezawa et al. (2018) were calculated for $\delta^{13}\text{C}-\text{CH}_4$ with NIWA as the reference laboratory and for $\delta\text{D}-\text{CH}_4$ with UU as the reference laboratory, and these were transferred to the MPI-BGC scale as explained in the main text.

Table 2: List of interlaboratory offsets calculated with MPI-BGC and TU/NIPR as the reference laboratory. Rows highlighted in grey have been indirectly calculated due to no overlap between the test and reference laboratory. Values marked with * are the modified offsets mentioned in Section 2.2.3.

Reference Laboratory	Test Laboratory	Tracer	Umezawa et al. (2018)		Nearest-neighbour approach		Smoothed data approach	
			Offset (‰)	Uncertainty (‰)	Offset (‰)	Uncertainty (‰)	Offset (‰)	Uncertainty (‰)
MPI-BGC	INSTAAR	$\delta^{13}\text{C}$	0.25	0.12	0.33	0.06	0.30	0.11
	NIWA	$\delta^{13}\text{C}$	0.12	0.12	0.23	0.07	0.21	0.03
	RHUL	$\delta^{13}\text{C}$	0.24	0.12	0.31	0.06	0.29	0.11
	TU/NIPR	$\delta^{13}\text{C}$	0.45	0.13	0.54	0.06	0.54	0.08
	UHEI	$\delta^{13}\text{C}$	0.08*	0.13	0.21	0.06	0.18	0.11
	UU	$\delta^{13}\text{C}$	0.08*	0.14	0.24	0.07	0.26	0.06
	INSTAAR	δD	-9.0	1.8	-11.9	0.7	-10.9	2.2
	TU/NIPR	δD	-8.9	1.3	-10.8	0.6	-10.8	1.6
	UHEI	δD	0.4	1.8	3.2	0.9	3.4	2.2
	UU	δD	4.2	1.2	2.1	0.5	2.4	1.6
TU/NIPR	INSTAAR	$\delta^{13}\text{C}$	-0.20	0.04	-0.19	0.06	-0.19	0.11
	MPI-BGC	$\delta^{13}\text{C}$	-0.45	0.13	-0.56	0.06	-0.55	0.09
	NIWA	$\delta^{13}\text{C}$	-0.33	0.04	-0.31	0.06	-0.31	0.03
	RHUL	$\delta^{13}\text{C}$	-0.21	0.05	-0.22	0.06	-0.24	0.11
	UHEI	$\delta^{13}\text{C}$	-0.37	0.06	-0.35	0.06	-0.37	0.06
	UU	$\delta^{13}\text{C}$	-0.37	0.08	-0.35	0.06	-0.34	0.06

	INSTAAR	δD	-0.1	1.4	-1.0	0.6	-0.1	1.8
	MPI-BGC	δD	8.9	1.3	10.9	0.6	10.8	1.3
	UHEI	δD	9.3	1.4	14.1	0.9	14.2	1.7
	UU	δD	13.1	0.6	12.9	0.5	13.0	1.6

3. Results

Fig. 3 and Table 2 show the overall offsets calculated for seven laboratories relative to MPI-BGC as a reference laboratory.

The offsets are calculated using the nearest-neighbour (red symbols) and smoothed data (blue symbols) approaches. The black-outlined white bars in the figure represent the offsets reported by Umezawa et al. (2018), rescaled to MPI-BGC as the reference laboratory, and including the modifications mentioned in section 2.2.3. The grey bars indicate the values of Umezawa et al. (2018) without these modifications. Thus, the differences between the grey and black-outlined bars in Fig. 3 and S1 are due to the revised UHEI vs. NIWA and UU vs. NIWA offsets.

The $\delta^{13}\text{C}$ - CH_4 offsets from the nearest-neighbour approach range from +0.2 to +0.5 ‰. TU/NIPR has the highest average offset relative to MPI-BGC, slightly exceeding previous estimates from Umezawa et al. (2018). INSTAAR, RHUL, UU, and NIWA follow, all showing slightly higher offsets than Umezawa's reported values. The offsets for UHEI are indirectly calculated due to a lack of direct overlap with MPI-BGC data. Overall, the offsets relative to the MPI-BGC scale determined in this study are about 0.1 ‰ higher than those from Umezawa et al. (2018) but remain within the reported instrumental uncertainty.

For δD - CH_4 , the derived scale offsets range from -11.9 ‰ to +3.2 ‰. TU/NIPR and INSTAAR show a negative offset of about 10 ‰, slightly lower than the value reported by Umezawa et al. (2018). UU also has a slightly more negative offset than previously reported. Due to the absence of direct comparisons, offsets for UHEI and INSTAAR are estimated indirectly via TU/NIPR. INSTAAR shows a more negative offset than Umezawa's findings (qualitatively consistent with INSTAAR and UU), while UHEI shows a higher offset.

Offsets derived from the smoothed data approach are generally consistent with those from the nearest-neighbour approach. The error bars are typically smaller for both methods than those reported by Umezawa et al. (2018) since a large number of measurements are compared. Thus, analytical errors cancel out to a large degree. Overall, the offsets calculated from the atmospheric air samples in this study agree with those reported by Umezawa et al. (2018). The modifications applied to the Umezawa et al. (2018) results (white bars relative to grey bars) enhance the compatibility between the offsets derived in this study and those reported by Umezawa et al. (2018) for all laboratories, particularly for UHEI vs. MPI-BGC.

3.1. Differences between the MPI-BGC and TU/NIPR as reference laboratories

When comparing the offsets between our determination and Umezawa et al. (2018), the agreement is slightly better for most laboratories when TU/NIPR is used as a reference laboratory (Fig. S1) than MPI-BGC (Fig. 3). For example, the $\delta^{13}\text{C}$ - CH_4 offset for INSTAAR relative to MPI-BGC is 0.25 ‰ in Umezawa et al. (2018) and 0.33 ‰ in our study, yielding a difference of 0.08 ‰. In contrast, when scaling to TU/NIPR, the offset for INSTAAR is -0.20 ‰ in Umezawa et al. (2018) and -0.19 ‰ in our study, resulting in a smaller difference of 0.01 ‰. This pattern holds across laboratories, and the mean offset difference for all laboratories is about 0.1 ‰ with MPI-BGC as the reference and 0.01 ‰ with TU/NIPR. The reduction of the relative scale difference between the two reference laboratories suggests that the central value of the MPI-BGC offset relative to NIWA might have been overestimated by approximately 0.1 ‰ in Umezawa et al. (2018); note that this is still in agreement with the uncertainty reported there. For δD - CH_4 , there are also slight differences when changing the scale from MPI-BGC to TU/NIPR. Specifically, INSTAAR's offset relative to MPI-BGC is -9.0 ‰ in Umezawa et al. (2018) and -11.9 ‰ in our study, leading to a difference of 2.9 ‰. However, when TU/NIPR is used as the reference, the INSTAAR offset is -

the agreement with Umezawa et al. (2018) is observed for UU; however, the difference becomes larger for the UHEI dataset. Lastly, the difference in scales for both reference laboratories with respect to Umezawa et al. (2018) is not systematic across laboratories. It is of the same order as the reported error bars.

3.2. Hemisphere-specific offsets

The nearest-neighbour approach may also be applied to hemisphere-specific datasets for each pair of laboratories. Given that the SH data points are fewer and show smaller seasonal variations, the interlaboratory offsets slightly differ between the two hemispheres. For example, the NH/SH mean $\delta^{13}\text{C-CH}_4$ offsets for INSTAAR, UU, and TU/NIPR relative to MPI-BGC are 0.34/0.33 ‰, 0.25/0.23 ‰, and 0.53/0.56 ‰, respectively. These values are very close to the global offset values of 0.33 ‰, 0.24 ‰, and 0.54 ‰, respectively. For $\delta\text{D-CH}_4$, the NH/SH mean offsets for TU/NIPR and UU relative to MPI-BGC are -10.7/-11.9 ‰ and 2.2/0.5 ‰, respectively, compared to the global offset values of -10.8 ‰ and 2.1 ‰.

3.3. Offsets derived from samples at the same station and offsets between similar high-latitude stations

To check the validity of our assumption of the high-latitude intra-hemispheric air masses being ‘well-mixed’, we performed two additional tests: (a) compare different laboratories measuring at the same station, and (b) compare different stations measured by the same laboratory. In the first case, we used Alert (NH) and Neumayer (SH) with the MPI-BGC as the reference laboratory. The interlaboratory offsets from measurements at Alert for $\delta^{13}\text{C-CH}_4$ are 0.34 ‰ (INSTAAR), 0.15 ‰ (UU), and 0.32 ‰ (RHUL). These values are consistent with the overall interlaboratory offsets when multiple stations measured by the same laboratory are grouped together, which are 0.33 ‰, 0.24 ‰, and 0.31 ‰, respectively. Similarly, for $\delta\text{D-CH}_4$, the station-specific offset for UU is 2.9 ‰, which compares well with the overall interlaboratory offset of 2.1 ‰ when multiple stations measured by the same laboratory are grouped together.

For the second case, we use INSTAAR, as it is the only laboratory measuring more than one station in the NH. The $\delta^{13}\text{C-CH}_4$ offsets w.r.t Barrow (longest sampling station) as the reference station are 0.07 ± 0.01 ‰ (Zeppelin), 0.07 ± 0.01 ‰ (Alert) and 0.13 ± 0.02 ‰ (Summit), and $\delta\text{D-CH}_4$ offsets w.r.t Barrow are 1.4 ± 0.13 ‰ (Zeppelin) and 0.9 ± 0.5 ‰ (Alert). These differences arise from the spatial variability in isotopic values taken at the four stations. The uncertainty is calculated as the standard deviation of isotopic measurements taken at the same time (7-day window); to capture how much the measurements vary with respect to each other due to differences in air masses at high latitudes, resulting in approximately 0.06 ‰ for $\delta^{13}\text{C-CH}_4$ and 0.5 ‰ for $\delta\text{D-CH}_4$. Such differences between stations suggest that the air measured at each station could be influenced by regional CH_4 sources, leading to small but measurable differences. Station Barrow, located amidst boreal wetlands, exhibits a more depleted $\delta^{13}\text{C-CH}_4$ signature than others, and this station is notoriously difficult to reproduce in global models (Basu et al., 2022). Excluding Barrow in our analysis brings the MPI-BGC vs INSTAAR $\delta^{13}\text{C-CH}_4$ offset from 0.33 ± 0.06 to 0.33 ± 0.04 ‰, which is not a significant change, and including Barrow ensures that the spatial variability, which is a necessary component of harmonising high-latitude datasets, is accounted for. This ensures that stations grouped by laboratory with spatial variability incorporated in their uncertainties are suitable for determining interlaboratory offsets from high-latitude atmospheric measurements.

Another approach is to filter out samples that do not represent the same well-mixed air by excluding measurements with significant differences in CH_4 mole fraction between the test and reference laboratory. We apply a stringent selection criterion to investigate this by retaining only nearest-neighbour pairs with CH_4 mole fraction differences of less than ± 10 ppb. This filtering yields nearly identical mean offset values for both isotopes, supporting the robustness of the derived offsets and indicating that differences between stations do not significantly influence the results.

Figure 4 shows the merged time series of all laboratories after accounting for laboratory-specific offsets (mean $\pm$ uncertainty) derived from the nearest-neighbour approach. The results from all laboratories capture very similar variability, including the reversal of $\delta^{13}\text{C-CH}_4$ and $\delta\text{D-CH}_4$ trends since 2010. The conversion of all datasets to a common reference scale (i.e. MPI-BGC: JRAS-M16) ensures a robust, offset-corrected dataset. Results using TU/NIPR as the reference scale are nearly identical and linearly transferable from MPI-BGC. Therefore, the harmonised dataset could be easily rescaled to either reference laboratories.

4. Discussion

4.1. Comparison of interlaboratory offsets to Umezawa et al. (2018)

Similar to Umezawa et al. (2018) and numerous smaller intercomparison activities, we find that interlaboratory offsets are larger than the measurement uncertainties from individual laboratories. This is due to the different approaches of individual laboratories to link the isotope composition of CH_4 to the international RMs, as described extensively in Umezawa et al. (2018). It is reassuring that the offsets reported in 2018 from a focused RR exercise using four gas cylinders that were shipped to participating laboratories agree within errors with the offsets determined from our approach using atmospheric air measurements up to 2023.

In their study, Umezawa et al. (2018) reviewed and discussed various offset values reported in previous research before recommending a specific offset value. As outlined in Section 2.2.4, we opted for different values in two instances, as these choices improved cross-laboratory consistency, although strong experimental evidence has yet to be provided. The first modification concerns the $\delta^{13}\text{C-CH}_4$ offset between the UHEI and NIWA laboratories (NIWA being the reference laboratory in Umezawa et al., 2018). Two previously reported values, -0.04 ± 0.04 ‰ (Poß, 2003; Behrens et al., 2008) and -0.07 ± 0.04 ‰ (Nisbet et al., 2005), are consistently around 0.1 ‰ higher than the value of -0.169 ± 0.031 ‰ reported by Umezawa et al. (2018) based on Levin et al. (2012). By adopting the value of -0.04 ‰, the offset for UHEI vs. MPI-BGC shifts from -0.05 ‰ to 0.08 ‰, ensuring that the differences between our results and those of Umezawa et al. (2018) relative to MPI-BGC are more consistent across laboratories (close to 0.1 ‰, Fig. 3). To examine this choice further, we calculated the NIWA vs. UHEI offset for SH (Arrival Heights and Neumayer stations) for the same period (1992-2008) considered in Levin et al. (2012) and obtained -0.11 ± 0.08 ‰ ensuring that the -0.169 ± 0.031 ‰ is in fact an overestimation of the interlaboratory offset.

For UU vs. NIWA, the interlaboratory offsets become more consistent if we use an offset of -0.08 ± 0.11 ‰ as reported by Umezawa et al. (2018) instead of -0.04 ± 0.03 ‰ from Schmitt et al. (2013) (this value was finally selected in the Umezawa et al. (2018) study). Choosing the value -0.08 ± 0.11 ‰ changes the MPI-BGC vs. NIWA offsets by $+0.04$ ‰, and since MPI-BGC is used as a reference laboratory, the values of all laboratories versus MPI-BGC change slightly. The changes due to these two adjustments are shown as the differences between the grey and black outlined bars in Fig. 3. Adapting these adjusted values leads to data consistency between the Umezawa et al. (2018) and this study, across all laboratories.

For $\delta\text{D-CH}_4$, the offsets computed agree with the ones reported in Umezawa et al. (2018) within the order of the instrumental uncertainty and the error reported in Umezawa et al. (2018). Note, however, that there is no direct overlap between UHEI and MPI-BGC for $\delta\text{D-CH}_4$, and the MPI-BGC offset is transferred via TU/NIPR as an intermediate laboratory, which, too, has limited overlap in the SH. The comparison of our results with Umezawa et al. (2018) suggests that the offsets between laboratories remain relatively constant over time. In reverse, this also supports using high-latitude ambient measurements for determining such offsets. For $\delta\text{D-CH}_4$, the differences are more pronounced, particularly for INSTAAR and UHEI, and are relatively greater (> 2 ‰) than those for UU and TU/NIPR. These discrepancies underscore the need for ongoing interlaboratory comparisons and adjustments to ensure the accuracy and traceability of methane isotopic measurements.

to spatial uncertainties between stations. The bootstrap method was chosen over standard deviation (SD) and standard error (SE) for estimating uncertainty because SD, as used by Umezawa et al. (2018) from various RR exercises, does not apply to our offsets. Our offsets are calculated from several hundred atmospheric measurements, but never from the same sample. SD assumes that variability comes from repeated measurements, but in this case, the offsets also stem from comparisons across different periods and sampling protocols. The SE, which reduces the uncertainty with the square root of the sample size, is also inappropriate here, as the offsets are based on different datasets from each laboratory, with unequal data point distributions per station. Bootstrap resampling directly estimates uncertainty by generating empirical confidence intervals from the data, capturing the underlying variability and providing more reliable results, particularly for datasets with non-normal distributions or uneven data distributions.

An advantage of the nearest-neighbour approach is that it can help detect temporally shifting offsets in the data harmonisation effort. In principle, this information could apply temporally shifting offsets to the individual datasets relative to one reference dataset. However, the risk of this procedure is that the reference laboratory is assumed to be free of temporal biases. If this is not the case, and all other laboratories are converted to this scale, potential biases in the reference laboratory would be emphasised and characterised as atmospheric variability. Analysing time-dependent offsets between laboratory time series may not completely eliminate biases when merging time series, but it does provide valuable insights into potential issues in different laboratories.

Maintaining a consistent internal calibration scale over long periods, often spanning decades, is challenging. Changes in analytical or data processing procedures can introduce bias into high-precision records, potentially leading to inaccurate conclusions about long-term isotope records. However, when laboratories change their data processing methods, they apply measures to ensure consistency and accuracy across the dataset. Comparisons of ongoing atmospheric sample results from similar locations across laboratories, in addition to RR and other interlaboratory intercomparison efforts, can help maintain long-term consistency in data records and provide a means to an early indication of inconsistencies. Continuous interlaboratory comparisons help ensure the accuracy and reliability of these measurements, and the nearest-neighbour approach is an easy diagnostic that can help to separate real natural variability from changes in laboratory scales. In recognition of the potential for laboratory offset variations, the WMO-GAW has a history of recommending continuous interlaboratory comparisons, including same-air measurements (RRs) and co-located sampling (WMO-GAW #292, 2024).

4.2. Challenges in harmonising atmospheric datasets

The assumption that high-latitude intra-hemispheric samples represent well-mixed air may not always hold, particularly in the NH, where local emissions and synoptic-scale atmospheric transport can create spatial differences in air masses. We find some inter-station variability in the high NH latitudes, possibly due to some local emissions in the NH. In contrast, the SH stations, surrounded by the ocean, tend to have more homogeneous air masses (Levin et al., 2012). These geographical differences can potentially contribute to intra-hemispheric offsets when comparing air sampled at different NH stations. Therefore, spatial variability has been incorporated into the uncertainty estimates in the nearest-neighbour approach. Variable sampling frequency may also impact interlaboratory comparisons, with some stations sampled every few days while others are sampled once a week or even once a month. In cases where there is no overlap between laboratories, offsets must be calculated indirectly, which adds uncertainty to the results.

Intra-hemispheric interlaboratory sample comparisons and interlaboratory RR exercises each offer unique advantages for harmonising datasets. While the comparison method facilitates offset estimates over extended measurement periods, they are susceptible to errors during sample collection and handling, as well as spatial variability due to atmospheric conditions. To address spatial variability more explicitly, a transport model could quantify the differences in CH₄ composition at various

425 reduce the uncertainty estimate, particularly since spatial variability is the largest contributor to the uncertainty budget in our method. However, this method would also introduce model-dependent uncertainties and is beyond the scope of the current study. RR exercises directly compare the same air sample across laboratories and provide a precise snapshot of interlaboratory differences by evaluating instrumental measurement conditions only and excluding sample handling or site-specific conditions.

When comparing the results of different long-term atmospheric measurement programs, it is ideal to compute offsets
430 between samples representing the same sampling date and measure them relatively soon after sampling by their respective laboratories. However, for some of our datasets (TU/NIPR, UU), isotope measurements were conducted a long time (up to three decades) after the sampling dates. One prerequisite is that the air stays stable in the sample containers. The good agreement between the measurements performed by UHEI and UU up to decades apart shows that this can be realised in high-pressure cylinders. However, in this case, the time series comparison does not allow for the detection of possible temporal
435 shifts in laboratory calibration scales.

The practicality of intra-hemispheric interlaboratory sample comparisons hinges on the availability of data. If the data are readily available, this method can be cost-effective. Our method applies to laboratories already performing continuous monitoring and comparison over extended periods, providing a consolidated understanding of temporal trends of the isotopic composition of CH₄ and correcting for potential biases introduced through air sampling, storage, transport, and handling. It
440 leverages existing infrastructure and data collection efforts, potentially reducing costs associated with the frequent shipping of cylinders as needed for RR exercises. Combining both approaches can provide a comprehensive understanding of interlaboratory offsets. RR exercises establish baseline consistency, and co-located, or intra-hemispheric, harmonised time series samples can help monitor the stability of these offsets over time.

4.3. Harmonised long-term datasets for high-latitude regions in both hemispheres.

445 Figure 4 shows the merged time series for CH₄, $\delta^{13}\text{C-CH}_4$ and $\delta\text{D-CH}_4$ at high northern and southern latitude stations after correcting for averaged offsets as described in 2.2.1. For consistency, CH₄ mole fraction records from the 8 laboratories were harmonised w.r.t MPI-BGC with the same nearest-neighbour approach. Most of the mole fraction data is obtained from the IRMS measurements for the corresponding isotope value, while some, as in the case of MPI-BGC, from a gas analyser (Jordan et al., 2024). The latter method has precise mole fraction sensors with much higher temporal resolution to follow CH₄
450 variations in more detail. The time series in both hemispheres exhibit clearly defined seasonality, supported by data from all participating laboratories. While averaged offsets are applied to account for interlaboratory differences, some laboratory-specific temporal variability remains visible. This is because individual data points may still reflect natural variability at their respective sites, and all points include measurement uncertainty. Correcting for offsets helps align the datasets, but does not eliminate the inherent noise present in the measurements. This highlights the importance of considering the corrected offsets
455 and the natural variability when interpreting the time series.

Nevertheless, since many data series are combined, possible artefacts from individual laboratories are more likely to average out, and the final data series should present a realistic time series of $\delta^{13}\text{C-CH}_4$ and $\delta\text{D-CH}_4$ over more than 30 years. We also produce a smoothed data series (Fig. 5) from this record that can be used directly in atmospheric box models. In the smoothed data, we exclude INSTAAR $\delta\text{D-CH}_4$ values from SH and pre-2015 MPI-BGC $\delta^{13}\text{C-CH}_4$ values as they may bring
460 additional artefacts to the short-term signal. Pre-2015 MPI-BGC $\delta^{13}\text{C-CH}_4$ values are excluded due to two main issues: (1) during the instrument setup at MPI-BGC between 2012 and 2014, early samples were analysed as 'single shots' instead of duplicates or triplicates and (2) cold traps in the MPI-BGC system were redesigned in October 2014, which improved CH₄ trapping efficiency and likely improved measurement quality compared to the earlier measurements. INSTAAR $\delta\text{D-CH}_4$ values from SH are excluded because they stand out from the other time series, even after offset correction (Fig. 4).

Deleted: a

monthly scales, with further improvements in comparability and sampling frequency expected to enhance temporal resolution. This work supports the call to establish a "central calibration laboratory" for CH₄ isotopes (similar to NOAA for CH₄ mole fraction and MPI-BGC for CO₂ isotopes) to ensure long-term traceability between RMs and atmospheric observations through common methane-in-air reference standards (Sperlich et al., 2016), as suggested in the Greenhouse Gases and Related Tracers Measurement Techniques (GGMT) recommendations, that would significantly enhance the achievable measurement compatibility (WMO-GAW #292, 2024).

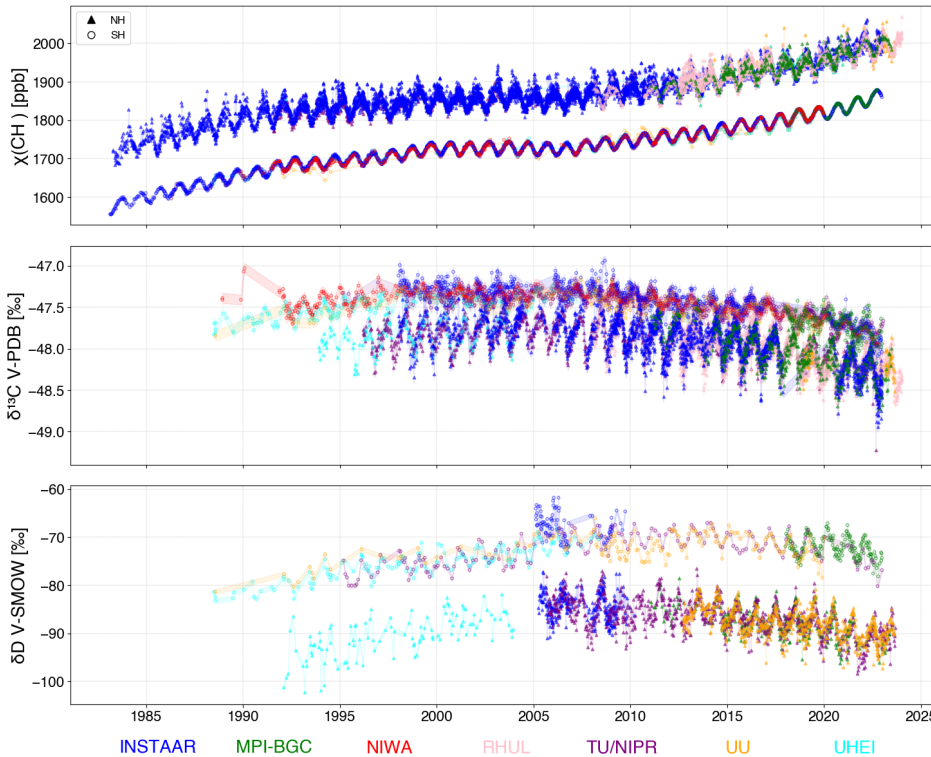


Figure 4: Harmonised data series with interlaboratory offset applied (mean $\pm$ uncertainty) to the observed data from eight laboratories using MPI-BGC as the reference laboratory. Relative to MPI-BGC, the TU/NIPR scale is enriched (+0.54 ‰) for $\delta^{13}\text{C}-\text{CH}_4$ and depleted (−10.8 ‰) for $\delta\text{D}-\text{CH}_4$.

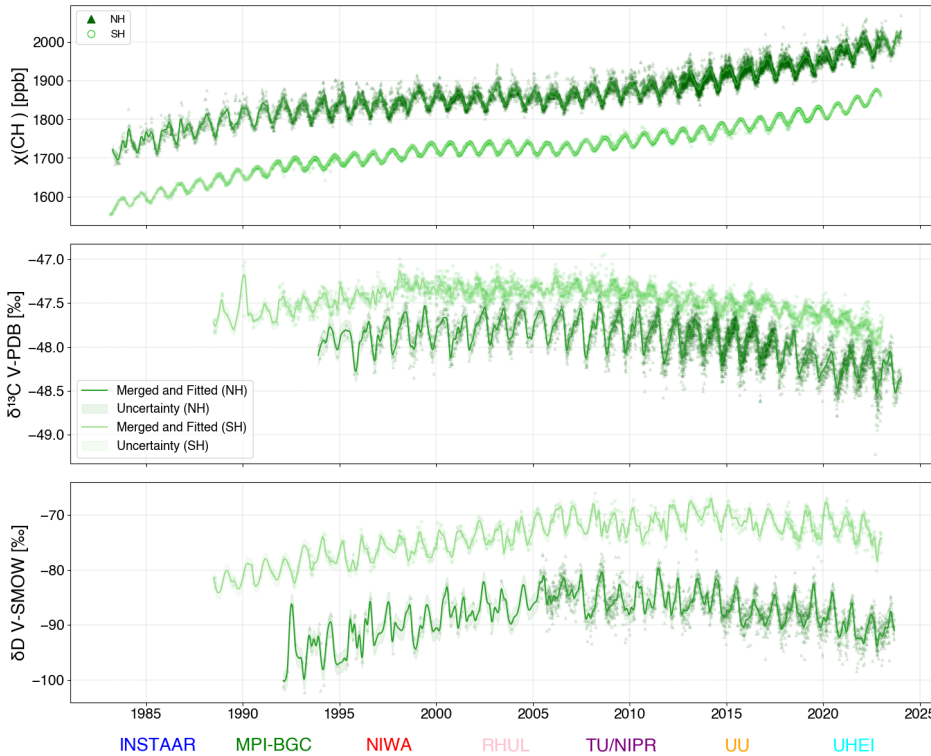


Figure 5: Merged and fitted data series with interlaboratory offset applied (mean $\pm$ uncertainty) to the observed data from eight laboratories using MPI-BGC as the reference laboratory. The NOAA CCGCRV algorithm is used for the fit, and the RMSE of residuals is plotted as the uncertainty band.

5. Concluding Remarks

The intra-hemispheric interlaboratory sample intercomparison method is a simple, cost-effective approach for identifying and quantifying measurement offsets between laboratories by leveraging long-term routine air samples. It provides robust estimates of relative differences to established reference scales such as MPI-BGC. This method harnesses thousands of ambient air samples and effectively captures harmonisation-related errors, which means that analytical uncertainty, as calculated with our Bootstrapping approach, will, to a large degree, average out. The largest contribution to the error is the real atmospheric variability of air between different sampling stations and dates. Additionally, offsets calculated by merging data from different stations agree with offsets calculated from a subset of data where only one station is considered (i.e., co-located intercomparison). In contrast, RR surveys use well-controlled, laboratory-designed experiments on identical samples, but with a limited number of data points measured during a specific period. Incorporating time series intercomparisons can complement deterministic approaches like RRs by offering extensive and continuous data. This helps confirm offset estimates, track instrument issues, and prevent shifts in calibration scales over time, ultimately improving the compatibility and interpretability of CH₄ isotope data.

One of the key results of our analysis is the consistency of the long-term trends in CH₄ isotopic composition from all the participating laboratories over multi-year periods. The offsets reported by Umezawa et al. (2018) were measured over different periods and within various subsets of CH₄ isotope-measuring laboratories. Despite this, the offsets calculated in our study,

these offsets are applied, atmospheric measurements spanning more than three decades align closely with their respective NH and SH merged time series curves (Fig. 4). This level of agreement is particularly noteworthy given the substantial effort, complexity, and logistical challenges involved in conducting methane-in-air isotope measurements.

Balancing the time series resolution with offset-related variability is crucial to maintaining accuracy when merging datasets. Using a 7-day window for synchronisation helps align datasets but risks excluding or reusing data points, highlighting the importance of considering both corrected offsets and natural variability in atmospheric analysis. As recommended in the World Meteorological Organisation Global Atmospheric Watch (WMO-GAW) report #292 (2024), we still remark on the importance of regular intercomparison of atmospheric measurements to assess measurement compatibility. In addition, by consistently comparing observation data periodically, laboratories can identify and correct possible discrepancies, leading to a better understanding of laboratory offsets and harmonised datasets.

Code/Data availability

The final harmonised dataset, as described in Figure 4, is available on the ICOS data portal (<https://doi.org/10.18160/V1Y4-NTK0>).

Author Contribution

BD, TR and MM conceptualised the manuscript. BD carried out the analysis, investigation, methodology and visualisation. MM helped with data collection, analysis, and visualisation. IL, HM, SEM, PS, RF, SM, TU, MM and TR helped with review and editing. All authors contributed equally towards the entry dataset.

Competing interests

Some authors are members of the editorial board of the journal Atmospheric Measurement Techniques.

Acknowledgements

The project has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States. Christine Groot Zwaafink received financial support from the ReGAME project funded by the Research Council of Norway (grant no. 325610). Part of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). We are grateful to the Norwegian Polar Institute's staff for their careful air sampling at Ny-Ålesund and to NIPR for their help with flask preparation and sample analysis. We are grateful to the staff of the Norwegian Polar Institute and the Japanese Antarctic Research Expedition for their careful air sampling at Ny-Ålesund and Syowa Station. A part of the observations by TU/NIPR are funded by JSPS KAKENHI Grant Numbers 15H03722, 20H01966, and 24K00702. This paper is dedicated to Ingeborg Levin, who started long-term methane isotope measurements in Europe.

Deleted: The project 21GRD04 isoMET project has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States

- 540 Assonov, S. S., and Brenninkmeijer, C. A. M.: On the O-17 correction for CO₂ mass spectrometric isotopic analysis, *Rapid Commun. Mass Spectrom.*, 17, 1007-1016, 2003.
- Ballantyne, A.P., Miller, J.B. and Tans, P.P., 2010. Apparent seasonal cycle in isotopic discrimination of carbon in the atmosphere and biosphere due to vapor pressure deficit. *Global Biogeochemical Cycles*, 24(3).
- 545 Basu, S., Lan, X., Dlugokencky, E., Michel, S., Schwietzke, S., et al.: Estimating emissions of methane consistent with atmospheric measurements of methane and $\delta^{13}\text{C}$ of methane, *Atmos. Chem. Phys.*, 22, 15351-15377, <https://doi.org/10.5194/acp-22-15351-2022>, 2022.
- Behrens, M., Schmitt, J., Richter, K.U., Bock, M., Richter, U.C., Levin, I. and Fischer, H., 2008. A gas chromatography/combustion/isotope ratio mass spectrometry system for high-precision $\delta^{13}\text{C}$ measurements of atmospheric methane extracted from ice core samples. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry*, 22(20), pp.3261-3269.
- 550 Bergamaschi, P., Schupp, M. and Harris, G.W., 1994. High-precision direct measurements of $^{13}\text{CH}_4/^{12}\text{CH}_4$ and $^{12}\text{CH}_3\text{D}/^{12}\text{CH}_4$ ratios in atmospheric methane sources by means of a long-path tunable diode laser absorption spectrometer. *Applied Optics*, 33(33), pp.7704-7716.
- Bousquet, P., Ciais, P., Miller, J. B., Dlugokencky, E. J., Hauglustaine, D. A., et al.: Contribution of anthropogenic and natural sources to atmospheric methane variability, *Nature*, 443, 439-443, 2006.
- 555 Brand, W.A., Rothe, M., Sperlich, P., Strube, M. and Wendeberg, M., 2016. Automated simultaneous measurement of the $\delta^{13}\text{C}$ and $\delta^2\text{H}$ values of methane and the $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values of carbon dioxide in flask air samples using a new multi cryo-trap/gas chromatography/isotope ratio mass spectrometry system. *Rapid Communications in Mass Spectrometry*, 30(13), pp.1523-1539.
- 560 Bréas, O., Guillou, C., Reniero, F. and Wada, E., 2001. The global methane cycle: isotopes and mixing ratios, sources and sinks. *Isotopes in Environmental and Health Studies*, 37(4), pp.257-379.
- Brenninkmeijer, C.A., Janssen, C., Kaiser, J., Röckmann, T., Rhee, T.S. and Assonov, S.S., 2003. Isotope effects in the chemistry of atmospheric trace compounds. *Chemical Reviews*, 103(12), pp.5125-5162.
- 565 Chandra, N., Patra, P.K., Fujita, R., Höglund-Isaksson, L., Umezawa, T., Goto, D., Morimoto, S., Vaughn, B.H. and Röckmann, T., 2024. Methane emissions decreased in fossil fuel exploitation and sustainably increased in microbial source sectors during 1990–2020. *Communications Earth & Environment*, 5(1), p.147.
- Dean, J.F., Middelburg, J.J., Röckmann, T., Aerts, R., Blauw, L.G., Egger, M., Jetten, M.S., de Jong, A.E., Meisel, O.H., Rasigraf, O. and Slomp, C.P., 2018. Methane feedbacks to the global climate system in a warmer world. *Reviews of Geophysics*, 56(1), pp.207-250.
- 570 Dlugokencky, E. J., Masarie, K. A., Tans, P. P., Conway, T. J., and Xiong, X.: Is the amplitude of the methane seasonal cycle changing?, *Atmos. Environ.*, 31, 21–26, [https://doi.org/10.1016/S1352-2310\(96\)00174-4](https://doi.org/10.1016/S1352-2310(96)00174-4), 1997.
- Eyer, S., Tuzson, B., Popa, M.E., Van Der Veen, C., Röckmann, T., Rothe, M., Brand, W.A., Fisher, R., Lowry, D., Nisbet, E.G. and Brennwald, M.S., 2016. Real-time analysis of $\delta^{13}\text{C}$ - and $\delta\text{D-CH}_4$ in ambient air with laser spectroscopy: method development and first intercomparison results. *Atmospheric Measurement Techniques*, 9(1), pp.263-280.
- 575 Fisher, R., Lowry, D., Wilkin, O., Sriskantharajah, S. and Nisbet, E.G., 2006. High-precision, automated stable isotope analysis of atmospheric methane and carbon dioxide using continuous-flow isotope-ratio mass spectrometry. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry*, 20(2), pp.200-208.
- 580 Fujita, R., Morimoto, S., Maksyutov, S., Kim, H.-S., Arshinov, G., Brailsford, M., Aoki, S., and Nakazawa, T.: Global and regional CH₄ emissions for 1995–2013 derived from atmospheric CH₄, $\delta^{13}\text{C-CH}_4$ and $\delta\text{D-CH}_4$ observations and a chemical transport model, *J. Geophys. Res.-Atmos.*, 125, e2020JD032903, <https://doi.org/10.1029/2020JD032903>, 2020.

- program of MPI-BGC, version 2022.1.
- 585 Keppeler, F., Laukenmann, S., Rinne, J., Heuwinkel, H., Greule, M., Whiticar, M. and Lelieveld, J., 2010. Measurements of $^{13}\text{C}/^{12}\text{C}$ methane from anaerobic digesters: Comparison of optical spectrometry with continuous-flow isotope ratio mass spectrometry. *Environmental science & technology*, 44(13), pp.5067-5073.
- Kirschke, S., Bousquet, P., Ciais, P., Saunio, M., Canadell, J.G., Dlugokencky, E.J., Bergamaschi, P., Bergmann, D., Blake, D.R., Bruhwiler, L. and Cameron-Smith, P., 2013. Three decades of global methane sources and sinks. *Nature geoscience*, 6(10), pp.813-823.
- 590 Lan, X., Basu, S., Schwietzke, S., Bruhwiler, L. M. P., Dlugokencky, E. J., et al.: Improved Constraints on Global Methane Emissions and Sinks Using $\delta^{13}\text{C}\text{-CH}_4$, *Global Biogeochem. Cycles*, 35, e2021GB007000, 10.1029/2021GB007000, 2021.
- Levin, I., Veidt, C., Vaughn, B.H., Brailsford, G., Bromley, T., Heinz, R., Lowe, D., Miller, J.B., Poß, C. and White, J.W.C., 2012. No inter-hemispheric $\delta^{13}\text{C}\text{CH}_4$ trend observed. *Nature*, 486(7404), pp.E3-E4.
- 595 Michel, S.E., Lan, X., Miller, J., Tans, P., Clark, J.R., Schaefer, H., Sperlich, P., Brailsford, G., Morimoto, S., Moossen, H. and Li, J., 2024. Rapid shift in methane carbon isotopes suggests microbial emissions drove record high atmospheric methane growth in 2020–2022. *Proceedings of the National Academy of Sciences*, 121(44), p.e2411212121.
- Miller, J.B. and Tans, P.P., 2003. Calculating isotopic fractionation from atmospheric measurements at various scales. *Tellus B: Chemical and Physical Meteorology*, 55(2), pp.207-214.
- 600 Monteil, G., Houweling, S., Dlugokencky, E. J., Maenhout, G., Vaughn, B. H., et al.: Interpreting methane variations in the past two decades using measurements of CH_4 mixing ratio and isotopic composition, *Atmos. Chem. Phys.*, 11, 9141-9153, Doi 10.5194/Acp-11-9141-2011, 2011.
- Morimoto, S., Aoki, S., Nakazawa, T., and Yamanouchi, T.: Temporal variations of the carbon isotopic ratio of atmospheric methane observed at Ny Ålesund, Svalbard from 1996 to 2004, *Geophys. Res. Lett.*, 33, L01807, <https://doi.org/10.1029/2005GL024648>, 2006.
- 605 Morimoto, S., Fujita, R., Aoki, S., Goto, D. and Nakazawa, T., 2017. Long-term variations of the mole fraction and carbon isotope ratio of atmospheric methane observed at Ny-Ålesund, Svalbard from 1996 to 2013. *Tellus B: Chemical and Physical Meteorology*, 69(1), p.1380497.
- Nisbet, E. G., Manning, M. R., Dlugokencky, E. J., Michel, S. E., Lan, X., et al.: Atmospheric Methane: Comparison Between Methane's Record in 2006–2022 and During Glacial Terminations, *Global Biogeochem. Cycles*, 37, 10.1029/2023gb007875, 2023.
- 610 Nisbet, E.G., Dlugokencky, E.J., Manning, M.R., Lowry, D., Fisher, R.E., France, J.L., Michel, S.E., Miller, J.B., White, J.W.C., Vaughn, B. and Bousquet, P., 2016. Rising atmospheric methane: 2007–2014 growth and isotopic shift. *Global Biogeochemical Cycles*, 30(9), pp.1356-1370.
- Parker, R. J., Boesch, H., McNorton, J., Comyn-Platt, E., Gloor, M., Wilson, C., Chipperfield, M. P., Hayman, G. D., and Bloom, A. A.: Evaluating year-to-year anomalies in tropical wetland methane emissions using satellite CH_4 observations, *Remote Sens. Environ.*, 211, 261–275, <https://doi.org/10.1016/j.rse.2018.02.011>, 2018.
- 615 Poß, C.: Investigation of the variability of atmospheric methane in polar regions based on trajectory analysis and the measurement of stable isotopes (in German), PhD thesis, Heidelberg University, 17–32, 2003.
- Press, W.H. and Rybicki, G.B., 1989. Fast algorithm for spectral analysis of unevenly sampled data. *Astrophysical Journal*, Part 1 (ISSN 0004-637X), vol. 338, March 1, 1989, p. 277-280., 338, pp.277-280.
- 620 Quay, P., Stutsman, J., Wilbur, D., Snover, A., Dlugokencky, E., et al.: The isotopic composition of atmospheric methane, *Global Biogeochem Cy*, 13, 445-461, 1999.

- Lowry, D., 2021. Boreas: a sample preparation-coupled laser spectrometer system for simultaneous high-precision in situ analysis of $\delta^{13}\text{C}$ and $\delta^2\text{H}$ from ambient air methane. *Analytical Chemistry*, 93(29), pp.10141-10151.
- Rigby, M., Montzka, S. A., Prinn, R. G., White, J. W. C., Young, D., et al.: Role of atmospheric oxidation in recent methane growth, *Proc. Natl. Acad. Sci. U. S. A.*, 114, 5373-5377, 10.1073/pnas.1616426114, 2017.
- Röckmann, T., Brass, M., Borchers, R., and Engel, A.: The isotopic composition of methane in the stratosphere: High-altitude balloon sample measurements, *Atm. Chem. Phys.*, 11, 13287-13304, 2011.
- Röckmann, T., Eyer, S., Van Der Veen, C., Popa, M.E., Tuzson, B., Monteil, G., Houweling, S., Harris, E., Brunner, D., Fischer, H. and Zazzeri, G., 2016. In situ observations of the isotopic composition of methane at the Cabauw tall tower site. *Atmospheric Chemistry and Physics*, 16(16), pp.10469-10487.
- Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J.G., Dlugokencky, E.J., Etiope, G., Bastviken, D., Houweling, S. and Janssens-Maenhout, G., 2016. The global methane budget: 2000–2012. *Earth System Science Data Discussions*, 2016, pp.1-79.
- Schaefer, H., Fletcher, S. E. M., Veidt, C., Lassey, K. R., Brailsford, G. W., et al.: A 21st-century shift from fossil-fuel to biogenic methane emissions indicated by (CH₄)-13C, *Science*, 352, 80-84, 10.1126/science.aad2705, 2016.
- Schmitt, J., Seth, B., Bock, M., Van der Veen, C., Möller, L., Sapart, C.J., Prokopiou, M., Sowers, T., Röckmann, T. and Fischer, H., 2013. On the interference of Kr during carbon isotope analysis of methane using continuous-flow combustion–isotope ratio mass spectrometry. *Atmospheric Measurement Techniques*, 6(5), pp.1425-1445.
- Schwietzke, S., Sherwood, O.A., Bruhwiler, L.M., Miller, J.B., Etiope, G., Dlugokencky, E.J., Michel, S.E., Arling, V.A., Vaughn, B.H., White, J.W. and Tans, P.P., 2016. Upward revision of global fossil fuel methane emissions based on isotope database. *Nature*, 538(7623), pp.88-91.
- Sperlrich, P., Moossen, H., Geilmann, H., Bury, S. J., Brown, J. C. S., et al.: A robust method for direct calibration of isotope ratios in gases against liquid/solid reference materials, including a laboratory comparison for $\delta^{13}\text{C}$ -CH₄, <https://doi.org/10.1002/rcm.8944>, 2020.
- Sperlrich, P., GuilleVIC, M., Buizert, C., Jenk, T. M., Sapart, C. J., et al.: A combustion setup to precisely reference delta C-13 and delta H-2 isotope ratios of pure CH₄ to produce isotope reference gases of delta C-13-CH₄ in synthetic air, *Atmospheric Measurement Techniques*, 5, 2227-2236, 10.5194/amt-5-2227-2012, 2012.
- Sperlrich, P., Uitslag, N. A. M., Richter, J. M., Rothe, M., Geilmann, H., et al.: Development and evaluation of a suite of isotope reference gases for methane in air, *Atmospheric Measurement Techniques*, 9, 3717-3737, 10.5194/amt-9-3717-2016, 2016.
- Steur, P.M., Botter, D., Scheeren, H.A., Moossen, H., Rothe, M. and Meijer, H.A., 2023. Preventing drift of oxygen isotopes of CO₂-in-air stored in glass sample flasks: new insights and recommendations. *Isotopes in Environmental and Health Studies*, 59(3), pp.309-326.
- Tanaka, M., Nakazawa, T., Shiobara, M., Ohshima, H., Aoki, S., Kawaguchi, S., et al.: Variations of atmospheric carbon dioxide concentration at Syowa Station (69°00', 39°35'), Antarctica, *Tellus B*, 39, 72–79, <https://doi.org/10.3402/tellusb.v39i1-2.15324>, 1987.
- Thanwerdas, J., Saunois, M., Berchet, A., Pison, I., and Bousquet, P.: Investigation of the renewed methane growth post-2007 with high-resolution 3-D variational inverse modeling and isotopic constraints, *Atmos. Chem. Phys.*, 24, 2129-2167, 10.5194/acp-24-2129-2024, 2024.
- Thoning, K.W., Tans, P.P. and Komhyr, W.D., 1989. Atmospheric carbon dioxide at Mauna Loa Observatory: 2. Analysis of the NOAA GMCC data, 1974–1985. *Journal of Geophysical Research: Atmospheres*, 94(D6), pp.8549-8565.

- 665 methane and hydroxyl, *Proc. Natl. Acad. Sci. U. S. A.*, 114, 5367-5372, 10.1073/pnas.1616020114, 2017.
- Umezawa, T., Brenninkmeijer, C.A., Röckmann, T., Van Der Veen, C., Tyler, S.C., Fujita, R., Morimoto, S., Aoki, S., Sowers, T., Schmitt, J. and Bock, M., 2018. Interlaboratory comparison of $\delta^{13}\text{C}$ and δD measurements of atmospheric CH_4 for combined use of data sets from different laboratories. *Atmospheric Measurement Techniques*, 11(2), pp.1207-1231.
- White, J. W. C., Vaughn, B. H., and Michel, S. E.: University of Colorado, Institute of Arctic and Alpine Research (INSTAAR), Stable Isotopic Composition of Atmospheric Methane (D/H) from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network, 2005–2009, Version: 2016-04-26, available at: ftp://aftp.cmdl.noaa.gov/data/trace_gases/ch4h2/flask/ (last access: 18 November 2016), 2016.
- Whiticar, M. J.: The Biogeochemical Methane Cycle, in: *Hydrocarbons, Oils and Lipids: Diversity, Origin, Chemistry and Fate*, 1-78, 2020.
- 675 Woolley Maisch, C.A., Fisher, R.E., France, J.L., Lowry, D., Lanoisellé, M., Bell, T.G., Forster, G., Manning, A.J., Michel, S.E., Ramsden, A.E. and Yang, M., 2023. Methane Source Attribution in the UK Using Multi-Year Records of CH_4 and $\delta^{13}\text{C}$. *Journal of Geophysical Research: Atmospheres*, 128(21), p.e2023JD039098.
- Worden, J. R., Bloom, A. A., Pandey, S., Jiang, Z., Worden, H. M., et al.: Reduced biomass burning emissions reconcile conflicting estimates of the post-2006 atmospheric methane budget, *Nat Commun*, 8, 10.1038/s41467-41017-02246-41460, 680 ARTN 2227, 2017.