

The Dasgupta et al. manuscript aims to examine interlaboratory offsets in atmospheric CH₄ stable isotope measurements. The authors choose the approach of comparing atmospheric measurements at high latitude sites over the more traditional round-robin interlaboratory exercise in which cylinders containing identical air are measured at each lab. The authors argue that their approach offers the advantage of many samples over a long time span, resulting in overall more representative estimates of interlab offsets. The authors compare the interlab offsets obtained in this way with a prior study (Umezawa et al 2018) that used the round-robin approach. Finally, the interlab offsets are used to place all CH₄ isotope data sets on the same scale to yield unified datasets for NH and SH high latitudes.

Major comments:

Evaluation and monitoring of inter-laboratory measurement offsets for CH₄ isotopes is important, as CH₄ isotopic measurements are a key constraint for the CH₄ budget. The approach of intercomparing measurements (versus a round-robin exercise) is justifiable and has strong merits; mainly the large number of available samples and comparison over multiple years. The main drawback of this approach is spatial and temporal variability of air masses (much more important in Arctic than Antarctic), and this is thoroughly considered and discussed in the manuscript. I agree that applying this approach to relatively long-term data sets would result in some of the spatiotemporal variability effectively averaging out to zero, although systematic offsets between sites would remain. The statistical methods used in uncertainty estimation seem appropriate.

I think it would have been better to apply this approach in parallel with a more traditional round-robin intercomparison. Both approaches have merits, it has been quite a few years since the last intercomparison, and there does seem to be a fairly consistent offset between the prior round-robin offset estimates (Umezawa et al) and those obtained with the new approach (Fig. 3). That said, I realize that doing a round-robin with multiple labs can take years and is expensive so this is not something that I am recommending as part of the revisions.

In my opinion, the type of intercomparison /merging of some published datasets that is performed here is something that would typically be presented as part of the methodology for a study of global CH₄ isotope trends / budget, rather than as a stand-alone paper. I would therefore recommend the following additions to the study to make it more publishable:

- Examine and include discussion of temporal trends in the inter-laboratory offsets (if any)
- Use the interlab offsets determined here to place all the CH₄ isotope datasets (from all stations, not just high latitudes) from the laboratories involved in the intercomparison

on the same measurement scale and publish that larger unified dataset as part of the study

Response: We agree with the reviewer that our method is not a stand-alone solution for computing interlaboratory offsets but rather a complementary approach to traditional round-robin intercomparisons. There is actually a round-robin exercise presently ongoing, which involves more laboratories than the ones that participate in the present work. We still think that our approach is very useful by itself, especially for laboratories that perform high-precision atmospheric background measurements of the isotopic composition.

Further, we had initially indeed planned this part as a data harmonisation step for an inversion study. As this first part grew successively larger, and after establishing the most appropriate approach within our consortium, we decided to split it off and present this as a method manuscript where the data harmonisation effort is described in more detail than what would be possible if it were combined with an inversion study. The second part, building on this dataset, uses the offset-corrected, harmonised high-latitude isotope time series of methane to perform a 2-box Bayesian inversion. That manuscript (to be submitted) is aimed at source and sink apportionment and an evaluation of the $\delta D-CH_4$ as an additional constraint on isotope-enabled atmospheric inversions in improving our understanding of the methane budget.

1. Temporal trends in the inter-laboratory offsets (if any): We agree that co-located measurements can be used to monitor possible shifts in inter-laboratory offsets, and our consortium is actually making active use of this opportunity. However, this is particularly difficult for high-latitude sites because the sampling date is often not close to the analysis date. In particular, in the southern hemisphere (Antarctica), air samples are in many cases only shipped once per year to the laboratories. In addition, the UU laboratory re-measured cylinders collected by UHEI up to more than two decades after they were sampled. Thus, a general application of this approach is not easy, and it requires dedicated knowledge of the analytical conditions in each specialist laboratory. Therefore, we decided not to include this part in the manuscript, which is targeted at a wider community and should provide updated offsets between laboratories that can also be used by others in a straightforward manner to combine datasets from different laboratories.
2. Interlaboratory offsets determined here to place all the CH_4 isotope datasets (from all stations, not just high latitudes): We chose to offset correct and harmonise only high latitude datasets. Methane isotopes show spatial gradients across the globe, and we choose only the high-latitude regions because they are more remote from the main sources and reasonably well mixed; thus, records from different stations should still represent relatively similar air. This is not true for lower latitude sites, which are more and differently affected by local and regional sources.

Minor comments:

I found the discussion of some individual offsets between pairs of labs a bit scattered and hard to follow – perhaps consolidate into another table?

Response: We have chosen two reference laboratories, MPI-BGC and TU/NIPR, to harmonise datasets, which we think is the more appropriate and thorough approach rather than pair-wise comparisons. Including numbers with still more reference laboratories would result in too much data and take away from the central narrative of the work.

Line 458: “INSTAAR dD-CH₄ values from SH are excluded because they stand out from the other time series, even after offset correction (Fig. 4a).”

There is no panel labeled “a” on Fig 4; also INSTAAR SH dD data are not shown at all on this figure

Response: Apologies, this should be just ‘Fig. 4’. The revised Fig. 4 now includes both INSTAAR SH dD and Pre-2015 MPI data.

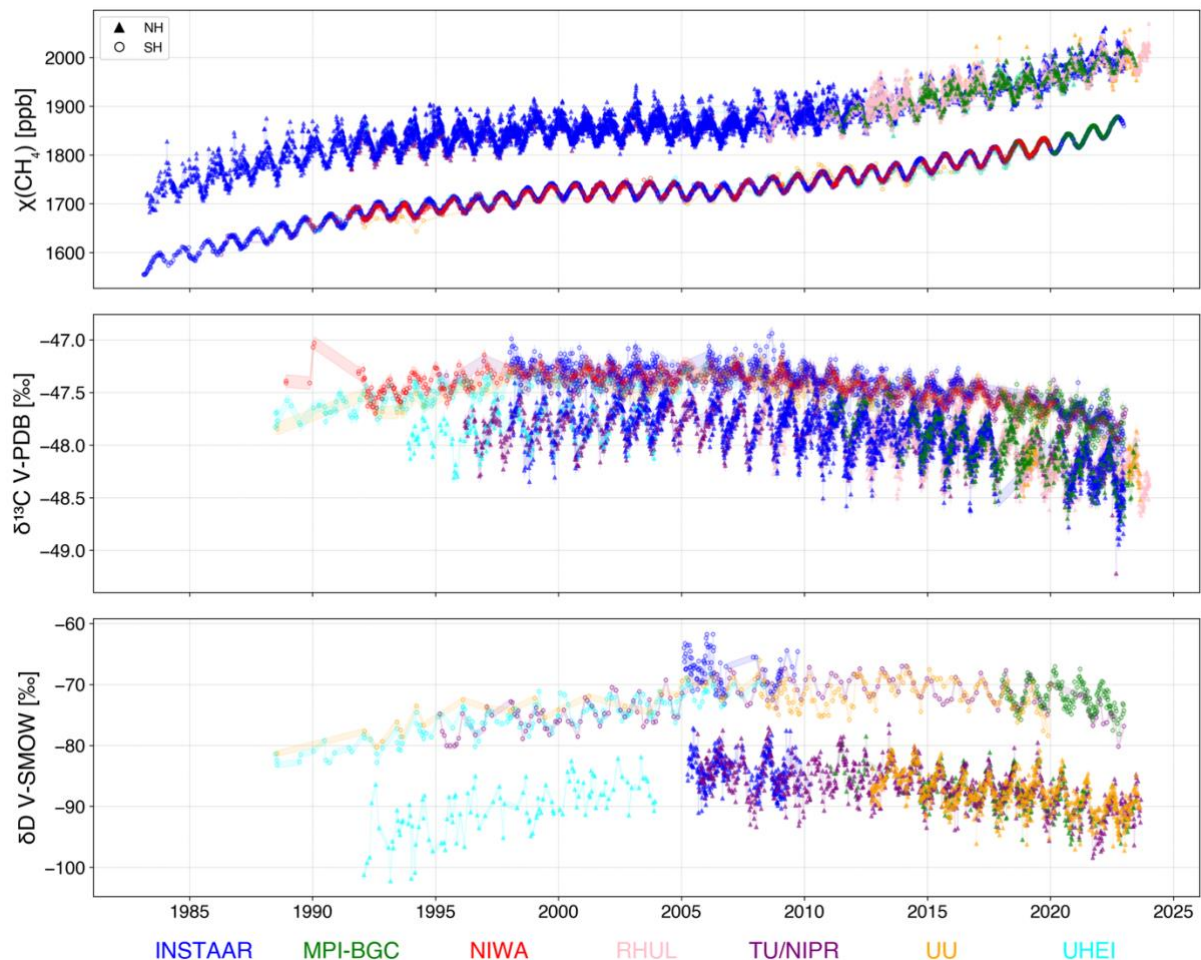


Figure 4: Harmonised data series with interlaboratory offset applied (mean $\pm$ uncertainty) to the observed data from eight laboratories using MPI-BGC as the reference laboratory. Relative to MPI-BGC, the TU/NIPR scale is enriched (+0.54 ‰) for $\delta^{13}\text{C}$ -CH₄ and depleted (-10.8 ‰) for δD -CH₄.

Line 241: “relative to NIWA” (remove “the”)

Response: Agreed.

The manuscript describes a method to quantify inter-laboratory offsets for measurements of $\delta^{13}\text{C}$ and δD in atmospheric methane, with the aim of harmonising each timeseries with a common reference. Offsets have been quantified previously by round-robin methods, and the method described here uses the timeseries of atmospheric measurements. After identifying periods of overlapping sampling, two approaches are presented to determine an offset relative to the selected reference laboratory MPI-BGC. These are a nearest neighbour approach that selects pairs of measurements sufficiently close in time then taking the difference and fitting the timeseries with a function, then taking the difference between pairs of functions. The offsets determined from the timeseries data are compared and found in agreement with previous intercomparison exercises (Umezawa *et al* 2018). Finally, the offsets are applied to the data to produce a harmonised timeseries for $\delta^{13}\text{C}$ and δD in the Northern and Southern Hemispheres.

Specific comments

Nearest-neighbour approach:

It was not clear to me how the pairs of test and reference data points are selected. Is the data for one laboratory from all sites combined, then any test point matched with the closest reference in time regardless of site? Or is the nearest-neighbour approach applied to the data from each site separately?

Response: The data from all high-latitude sites per laboratory and hemisphere are first combined into a lab-specific NH or SH dataset. Then each point from the reference lab is matched with the closest test lab data point in time. We apply this approach at the laboratory level instead of comparing site-by-site because it allows us to use larger and longer datasets. To account for possible differences between locations, we include a fixed spatial uncertainty in our calculations. We have clarified this in the revised manuscript.

Line 195 – Is the same value of spatial uncertainty used when two measurements at the same station are compared? For example, Table 1 shows that INSTAAR and MPI-BGC have overlapping records at Alert, so the comparison here uses measurements of the same air masses, while the other station pairings are geographically separated.

Response: The spatial uncertainty is a fixed value (0.06 ‰ for $\delta^{13}\text{C}\text{-CH}_4$ and 0.5 ‰ for $\delta\text{D}\text{-CH}_4$) based on differences between INSTAAR's northern stations (which do not include co-located or same station data), and it represents typical variability across high-latitude sites. Since it reflects broader atmospheric variability, it is not adjusted for co-located comparisons. This approach provides a conservative and consistent estimate of atmospheric variability across the entire dataset.

I suggest moving the two paragraphs from lines 378 and 388 from p. 13 into this section to describe the uncertainty calculation method.

Response: As the method applied in this paper is also part of its discussion, we have continued the points in lines 378-388 in the discussion in section 4.

Smoothed data approach:

Line 221 describes that the difference between two sites using the smoothed data approach is calculated from the difference between the long-term average of two sites. How does this account for the trend in the isotope ratio (i.e. the polynomial part)? That is, the mean of a longer running site will be earlier in the time series than a more recent site. Why not use the function to generate evenly spaced values for both sites then take the difference in the same way as the time-matching approach?

Response: For each laboratory pair, we first apply the NOAA CCGCRV smoothing to generate evenly spaced curve points over the entire record. We then restrict each smoothed series to its common overlap period and compute the mean of those matched points; the difference between those two means is the offset. Because the curves are defined by the same number of evenly spaced points across the same dates, this intrinsically accounts for both the long-term polynomial trend and the seasonal harmonics.

Line 228 – the authors recognise that the fitted curve should not be used to interpolate “significant gaps”. Can they provide a maximum duration that is used to exclude such portions.

Response: We exclude any overlapping segment shorter than one year or containing gaps longer than six months, since the NOAA fit becomes unreliable over very sparse data. In practice, this only affected the UU vs. MPI-BGC $\delta^{13}\text{C}$ -CH₄ comparison in the NH (2019–2022), which we have now noted explicitly in the revised text. By enforcing these minimum duration and maximum-gap criteria, we ensure that all smoothed-curve comparisons rest on robust, well-constrained fits.

Line 228 – “In this case, the extrapolated parts of the curve fit ...” should be corrected to “In this case, the interpolated parts of the curve fit ...”

Response: Agreed.

Line 231 – the uncertainty calculation is unclear. In the description “average sum of squares of the root mean square error” what is being averaged here? The calculation of the offset was described on line 221 as the difference between the long-term average for both sites, which implies a single value. If the average is over all pairings of test site and reference site then should there also be a spatial uncertainty that accounts for co-located and geographically separate sites?

Response: In the smoothed-data approach, we estimate uncertainty purely from how well each laboratory’s smoothed curve matches its own measurements. We calculate the root-mean-square error (RMSE) of the fit for each lab, that is, the typical deviation of the fitted curve points from the observed values, and then combine those two RMSEs into a single uncertainty by averaging their squared values and taking the square root of this average. Because this method compares two continuous, evenly spaced curves rather than individual station measurements, it does not include a separate spatial-uncertainty term: any geographic variability is inherently captured in the residuals used to compute each curve’s RMSE.

Harmonised long-term datasets

Line 459 - Can the authors provide more discussion for the statement “INSTAAR $\delta\text{D-CH}_4$ values from SH are excluded because they stand out from the other time series, even after offset correction”. It is not clear what is standing out for these measurements, and does it show some conditions where the offset correction is ineffective? Section 3.2 shows little difference between the hemisphere-specific offsets. The reference to Fig. 4a also needs to be updated as the figure does not have sub-figures with letter labels.

Response: We found that, even after applying the calculated lab-wide offset, some SH INSTAAR records lie systematically $\sim 5\text{‰}$ heavier than other SH points around that time (included now in the revised Fig.4, also pasted above). Because we impose a single, uniform offset per laboratory, the SH values remain biased, and it is difficult to filter out outliers. To prevent these residual artefacts from distorting the merged, smoothed time series, we have omitted the SH INSTAAR $\delta\text{D-CH}_4$ entirely from the harmonised plot (Fig. 5). However, the complete offset-corrected INSTAAR SH dataset is included in the ICOS data portal for any users who wish to include it.

Section 3.2 does not report INSTAAR $\delta\text{D-CH}_4$ hemisphere-specific offsets (no overlap with MPI-BGC), only TU/NIPR and UU.

Apologies, this should be just ‘Fig. 4’.

Figure 5 – The line colours for Northern Hemisphere and Southern hemisphere are very difficult to distinguish, can this be replotted more clearly.

Response: The line colours for NH and SH in Fig.5 have been updated.

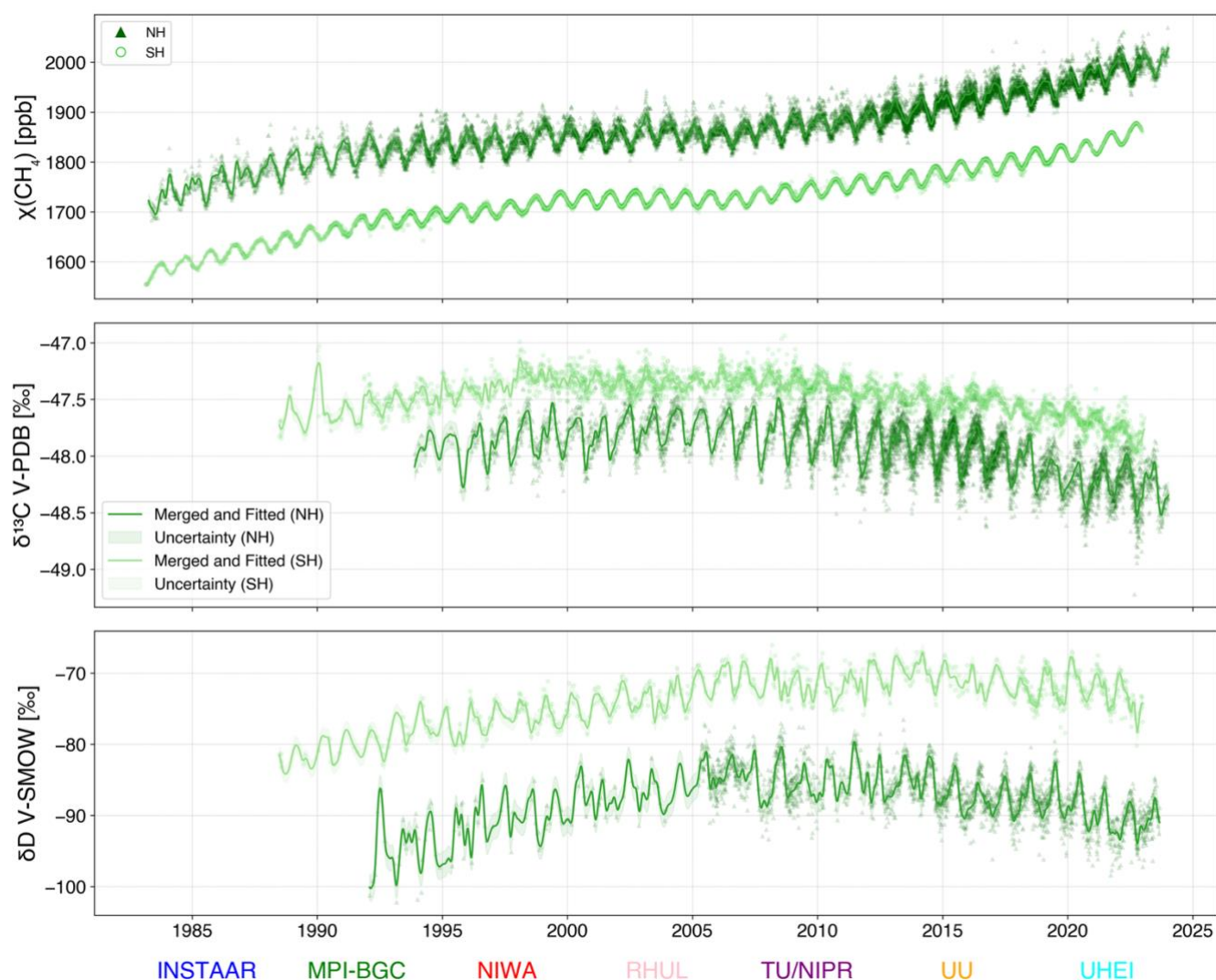


Figure 5: Merged and fitted data series with interlaboratory offset applied (mean $\pm$ uncertainty) to the observed data from eight laboratories using MPI-BGC as the reference laboratory. The NOAA CCGCRV algorithm is used for the fit, and the RMSE of residuals is plotted as the uncertainty band.

List of all relevant changes made in the manuscript:

1. Line 145 – Clarified in the Methods that in the nearest-neighbour approach, data from all high-latitude sites per laboratory and hemisphere are first combined into lab-specific NH or SH datasets before matching with the reference laboratory.
2. Line 228 – Added exclusion criterion for the smoothed-data approach: overlapping periods shorter than one year or containing gaps longer than six months are excluded.
3. Line 228 – Corrected “extrapolated parts of the curve fit” to “interpolated parts of the curve fit.”
4. Line 241 – Corrected phrasing from “relative to the NIWA” to “relative to NIWA.”
5. Line 458 – Corrected reference from “Fig. 4a” to “Fig. 4.”
6. Figure 4 – Updated to include both INSTAAR SH δD data and pre-2015 MPI δD -CH₄ data.
7. Figure 5 – Updated with clearer, more distinguishable colours for Northern Hemisphere and Southern Hemisphere lines.
8. Author contribution and Acknowledgement revised.