

Responses to individual reviewer comments (black text) are given below, in red text.

Summary:

This is a more impactful paper than the current title implies (see specific comments below). The study describes sources of systematic error affecting methane flux estimation using the tracer dispersion method, with acetylene as the tracer. It considers water vapour and other sources of spectroscopic/retrieval error, but the true novelty in the paper is most captured by the potential flux error that arises from non-calibration of acetylene analysers when used as a proxy for CH₄. While the study is specific to the Picarro C2H₂ analyser, the general need remains the same. The paper arrives at one flux bias number (7.6% - I think there's more to offer than this though), which is not insignificant in CH₄ emissions quantification methods. Calibration such as that described here is rarely used by those using tracer methods for survey work, but clearly should be.

We thank the reviewer for taking the time to review our manuscript and for their valuable comments, which have helped to make significant improvements to the manuscript.

The paper is an excellent fit to AMT and its readers. The paper has value to scientists and industry using Picarro (and other IR) analysers and those following methane emissions quantification methods and datasets. It is a well-presented, rigorous, and detailed study. While it is refreshing to see such a detailed, lengthy, and complete technical study such as this, some of the important conclusions of the paper get lost at present (see specific comments).

We are grateful to the reviewer for their confidence in our research. We appreciate the time the reviewer has taken to understand the fundamental outcomes of this research and its overall importance. We have made every effort to address their concerns. We agree that the emphasis of key outcomes could be improved, which we address through several modifications to the manuscript, as discussed in the responses below.

I wholeheartedly recommend publication, and the comments below are constructive in the hope that the potential of the study can be better realised. I also have some technical questions which may need to be addressed satisfactorily.

We thank the reviewer for their kind recommendation for publication. We hope we have addressed all of the reviewer's questions in the responses below.

Specific comments

1. Title: When I first saw the title, I thought this paper was about the simple calibration of an analyser. It is a fair bit more than this. It is about how calibration of an instrument significantly affects flux calculations. I would recommend a better title to attract readers to this important fact. It would be good to be explicit in the title that this is about impact on accuracy of fluxes.

We fully agree with the viewpoint of the reviewer. We have renamed the article to "Accuracy of tracer-based methane flux quantification: underlying impact of calibrating acetylene measurements". We hope this now emphasises the implications of failing to calibrate tracer gas measurements on methane fluxes.

2. Introduction: I'm really not sure the historical discussion of the discovery of the greenhouse effect is needed (with very old references in places). It is not explicitly

linked to the content of the paper. Also, CH₄ is the second-most important GHG according to the IPCC, after CO₂ (not the third) – please correct, or otherwise clarify. If you are referring to GWP rather than RF (or something else), please be clear. Line 42 cites Dlugokencky *et al.*, 2011 – please use the most recent reference for this regular review from Dlugokencky).

The reviewer is correct to state that methane is the second-most important anthropogenic greenhouse gas, according to the IPCC. We now explicitly clarify that the original statement refers to greenhouse gases in terms of overall radiative forcing. We have included a more recent reference to Schmidt *et al.* (2010) to reflect this. We thank the reviewer for this advice, to avoid ambiguity.

We recognise that some of the references are old. We have now added some more recent references to this introductory paragraph. For example, another citation to the latest IPCC report is made when discussing the effective radiative forcing of methane. In addition, we have now cited Saunio *et al.* (2024), which was not accepted for publication prior to submission of this manuscript. While we have included some of the latest publications to bring this section up to date, we would also prefer to maintain some older references to emphasise the prolonged nature of the issue of quantifying methane emissions. For example, in the same location as our citation to Dlugokencky *et al.* (2011), we also cited a more recent work from Lan *et al.* (2021), in the original manuscript. In this specific example there is a benefit to include references from different dates as this shows that uncertainties in the global methane budget has been a long and persistent problem. Hence, this fuels the motivation for the remainder of the manuscript.

Regarding the request to include the latest citation from Dlugokencky, it is not clear to us which specific work the reviewer is referring to here. We have already included the review from Lan *et al.* (2021). We have now included two additional review papers from the year 2019 and 2021 by Nisbet *et al.*, with Dlugokencky as a co-author in both articles. Perhaps, our citation to Saunio *et al.* (2024) can also satisfy this requirement, with Dlugokencky as a co-author in this work as well. We hope these are sufficiently recent citations to co-authored works by Dlugokencky, but we are unable to find a more recent paper by Dlugokencky as first author.

3. Line 100 – N₂O has a lifetime >100 years (greater than CH₄). The sentence here implies that use of N₂O may be less problematic than CH₄ (calling N₂O “finite lifetime”). All gases have a finite lifetime in any case. Please clarify and correct this misleading sentence.

The reviewer makes a very valid point here and we agree that this could be misleading. We have now provided the duration of the nitrous oxide lifetime in the text and removed the word “finite”.

4. Acetylene calibration: Ok.... So C₂H₂ standards aren't readily available. I understand that. Instead, an empirical calibration of mass flow controllers was used to produce a diluted air matrix. To do this, certified CH₄ standards were used to define the dilution characteristics of specific mass flow controllers. This sounds ok, but I do have some questions. 1/ Was the same MFC tested repeatedly (e.g. on different days, after switching on and off etc), and were the results always consistent? If not, can you be confident that there was no drift in the MFC rates when switching between CH₄ and C₂H₂? 2/ Can you be certain that there are no biases between these 2 gases in terms of

the experimental setup, e.g. C₂H₂ may be more sticky than CH₄ in sample lines – was flushing and equilibrium time used, or can you be sure there are no residual effects? 3/ When you take into account the precision of the CH₄ standard (i.e. how well known the concentration of the CH₄ reference cylinder is), how does that precision manifest as a percentage of the concentration of acetylene after dilution? And is that percentage similar to the gain factor calculated for C₂H₂? It would be worth including a discussion of this in the paper for completeness and the avoidance of doubt.

In response to the first concern, the behaviour of mass-flow controllers has indeed been observed to change with time, as stated in the manuscript. For this reason, absolutely no changes were made to any of the flow connections, except for swapping the methane calibration cylinder with the acetylene calibration cylinder. The mass-flow controllers were not powered off for the full duration of each calibration test (during both methane and acetylene enhancements). We now emphasise these points further in Section 2.4. We are confident that the mass-flow controllers did not drift during each calibration test as the same methane gas blend was sampled three times, with each blend sampled 5.75 hours apart. The average standard deviation between each of the three [CH₄] averages measured by the Picarro G2401 was ± 2 ppb (see end of Section 2.3), with [CH₄] reaching up to 12 ppm. Both methane and acetylene enhancements were measured within a 48-hour window as previously stated in Section 2.4, making it unlikely for any significant drift to occur during this period, considering the null drift observed over 11.5 hours.

The reviewer raises an astute point concerning the effect of the laboratory set-up on acetylene. The majority of gas during testing passed through Synflex 1300 tubing. Therefore, we have now conducted an additional [C₂H₂]_r test by alternating the same gas stream, passing through either pure stainless steel or Synflex 1300. The average measured [C₂H₂]_r was compared between the two gas streams and found to be identical (within uncertainty). This therefore confirms that Synflex 1300 has no noticeable effect on [C₂H₂], compared to stainless steel. It is highly unlikely that Synflex 1300 and stainless steel have the exact same impact on acetylene. It can therefore be assumed that Synflex 1300 has no net effect on acetylene, allowing for its use during laboratory testing. The results of this new analysis are presented in Section S2 in the Supplement, with the outcome of the analysis summarised at the start of Section 2.1, when presenting the use of Synflex 1300 for the first time. We thank the reviewer for motivating us to include this very useful analysis.

The final point is an interesting one, regarding the declared [CH₄] level in the methane calibration cylinder. We thank the reviewer for noticing this, due to their robust review of our methods. As this [CH₄] level could not be accurately measured for verification (due to the argon content of this cylinder), all [C₂H₂]_r (and hence all C_{MFC}) values rely on the accuracy of the declared [CH₄] level of the methane calibration cylinder, which has an uncertainty of $\pm 0.5\%$. Therefore, a new sensitivity analysis was conducted on the importance of this uncertainty range; a range of C_{MFC} values was calculated using the range in possible [C₂H₂]_r values (at each MFC setting), taking into account the uncertainty range in the declared [CH₄] of the methane calibration cylinder. Rather than simply testing the effect of the variation in C_{MFC} on deriving [C₂H₂] standards (used to calibrate Picarro G2203 [C₂H₂]_r measurements), we decided to also include the effect of the $\pm 3\%$ uncertainty associated with the declared [C₂H₂] level of the acetylene calibration cylinder. This yielded a range in [C₂H₂]_r levels (at each MFC setting) which were combined with corresponding C_{MFC} values to obtain [C₂H₂] reference standards. This analysis is presented in Section S5 in the Supplement, where Picarro G2203 [C₂H₂]_r calibration fits are presented by both maximising and minimising [C₂H₂] standards. This results in a possible extreme gain factor range of between 0.911 and 0.977,

which is presented in Section 2.4. It is important to note that this means that even in a worst-case scenario, the acetylene calibration gain factor is still significantly less than 1. Therefore, an acetylene calibration remains essential.

5. Flux error: A single number is given from the field trials of 7.6%. But a percentage isn't really that useful to those reading the paper, as a percentage of the flux will depend on all the factors you list that affect the quality of tracer release experiments and fluxes derived from them. A reader taking away the 7.6% number and using it as a guide to the effect of calibration would not be remembering something useful. However, I think you have all that you need here to derive a range of biases for different CH₄ fluxes and acetylene release rates. They're all linearly scalable. So, rather than 7.6%, would it not be better to include a table in the paper that describes CH₄ mass flux (g/s) error (and percent error), for a range of different assumed CH₄ emission fluxes, or even simply C₂H₂/CH₄ concentration ratios? This would be far more meaningful, and a better guide to others assessing how important calibration may be to their surveys. If you don't take up this recommendation, the discussion needs to be very explicit that the 7.6% number is not indicative of mass flux error due to instrument calibration and is unique to the conditions of this study (reducing the paper's impact).

We feel that the reviewer reached a different conclusion regarding the significance of this bias value, compared to the actual message we intended to convey. Perhaps this is because the original version of the manuscript lacked some discussion on the significance and origin of the bias value, which we now seek to rectify. This 7.6% value represents a bias that is largely independent of the sampling circumstances themselves including issues with sensor placement or downwind sampling location. Meanwhile, there is a $\pm 56\%$ random uncertainty between flux estimates which is unrelated to the acetylene calibration. This random uncertainty is indeed specific to the sampling campaign, which is discussed in depth in Section 4.2. However, this $\pm 56\%$ value is not the focus of this paper.

The 7.6% bias emphasised in the abstract and conclusion is principally due to the acetylene calibration. An acetylene calibration gain factor of 0.94 should result in a purely theoretical flux bias using uncalibrated measurements of exactly 6%, following eq. (4). However, this number is not exactly 6% for a number of reasons. The calibration was not valid below 1.19 ppb, meaning that all such [C₂H₂]_r measurements were fixed to 0 ppb. To account for this, a threshold methane mole fraction was also used, below which all [CH₄] measurements were set to the background methane mole fraction. Finally, the calibrated [C₂H₂]_r data included a small linear offset which is only valid above 1.38 ppb [C₂H₂]_r. These effects collectively cause flux bias to vary from exactly 6%.

The reviewer is therefore technically correct to state that this 7.6% bias is specific to this study. A higher number of [C₂H₂]_r measurements below 1.19 ppb from different transects will result in more [C₂H₂]_r values being fixed to 0 ppb. Alongside this, the specific location and magnitude of [CH₄] and [C₂H₂]_r measurements from each transect effects the [CH₄] threshold, designed to correct for the aforementioned effect. Hence, the magnitude of flux underestimation quantified for each transect using uncalibrated measurements instead of calibrated measurements varies. But by taking an average of fourteen different transects, individual variability in flux underestimation should average out. Therefore, although this 7.6% bias is specific to this study, we still believe it has merit, as it illustrates a typical order of magnitude level of flux bias induced by using uncalibrated measurements from a real-world sampling campaign, using the Picarro G2203 for tracer-based methane fluxes.

To include these points in the manuscript, additional extensive discussion has been included towards the end of Section 4.2. We hope this clarifies any doubt regarding the significance of this 7.6% flux underestimation using uncalibrated $[\text{C}_2\text{H}_2]_r$ measurements. We thank the reviewer for raising this, thus enabling us to thoroughly clarify this point.

6. There are a lot of important points to take from this paper, but they get lost in the extremely thorough discussion (and the conclusions). I would recommend completely changing the conclusions section to avoid unnecessary repetition of methods and instead focus on the salient conclusions about water vapour, need for acetylene calibration, and potential for flux error. It may also be useful to contextualise the error potential in terms of the magnitude of error from other flux methods (i.e. is it a comparable error if this calibration is ignored?). The important aspects of this useful paper will otherwise get missed.

We thank the reviewer for recognising the useful points in this paper. We agree that the key points of the paper may get lost in the conclusion in its original form. We have therefore rewritten the majority of the conclusion section and shortened it to place greater emphasis on the key outcomes of this work, with less focus on the methods. For example, details of the acetylene flow meter have now been omitted, with only the $Q_{\text{acetylene}}$ provided.

We have removed any discussion from Section 5 on the cause of the $\pm 56\%$ random uncertainty as it is a distraction and not the main focus of this work. Issues in the tracer flux method itself are also not the key outcome of this study. We thank the reviewer for their feedback, allowing us to place greater emphasis on the importance of the 8% bias, which is the main result, induced by failure to calibrate the Picarro G2203.

Regarding the suggestion to compare different flux uncertainties, in our opinion, the 8% uncertainty bias presented in this manuscript is not directly comparable to random uncertainties from other flux methods; a bias systematically forces the flux in one direction whereas a random uncertainty averages out over a sufficient number of samples. Furthermore, comparing flux errors from various flux methods is not the objective of this study, which will be the subject of a forthcoming study, as previously stated at the start of Section 4.1.

Technical comments:

1/ Line 152 – Full stop after bracket.

We thank the reviewer for spotting this oversight, which we have now rectified.